Submitted to the Annals of Statistics
arXiv: math.PR/0000000

SUPPLEMENTAL MATERIAL TO "PARTIALLY LINEAR ADDITIVE QUANTILE REGRESSION IN ULTRA-HIGH DIMENSION"

By Ben Sherwood and Lan Wang*

Johns Hopkins University and University of Minnesota

APPENDIX A: ADDITIONAL NUMERICAL RESULTS

The tables of the appendix provide additional numerical results. Table 1 summarizes simulation results for Q-SCAD, LS-SCAD, Q-MCP, LS-MCP with sample sizes 50, 100 and 200 for modeling the 0.7 conditional quantile for the heteroscedastic error setting described in Section 4 of the main paper. The MCP approaches, Q-MCP and LS-MCP, are the equivalent of Q-SCAD and LS-SCAD with the SCAD penalty function replaced by the MCP penalty function [Zhang (2010)]. Table 2 reports the simulation results for the MCP penalty function for the four simulation settings presented in Section 4 of the main paper. Table 3 is an extension of Table 3 from the main text with results for $p = 1200$ and 2400. Tables 4 and 5 report numerical results with the MCP penalty function for the real data analysis. Table 6 presents the simulation results using the MCP penalty function for simultaneous variable selection at different quantiles. The selection of the tuning parameter λ for the MCP penalty function uses the modified high-dimensional BIC criterion as we recommended for the SCAD penalty function. The tuning parameter a for MCP is set to 3, which is used as the default in the R package *ncvreg* for the least squares implementation of the MCP penalty [Breheny and Lee (2015)]. The results with MCP are observed to be similar as those with SCAD.

APPENDIX B: SOME TECHNICAL DERIVATIONS

B.1. Some details on the use of the theoretically centered basis functions. For the purpose of a self-contained proof, we provide below a brief derivation of the relationship between the estimators obtained using the theoretically (infeasible) centered basis functions and the estimators obtained using the original (feasible) basis function. The derivation essentially follows along the lines of Xue and Yang (2006).

*The work is supported in part by National Science Foundation grant DMS-1308960

Method	n	p	FV	TV	True	P	AADE	MSE
Q-MCP	50	100	9.40	3.33	0.00	0.25	0.69	2.56
Q-MCP	100	100	1.85	4.23	0.09	0.36	0.43	0.41
Q-MCP	200	100	0.42	4.70	0.48	0.71	0.23	0.09
Q-SCAD	50	100	7.08	3.34	0.00	0.31	0.64	2.28
Q-SCAD	100	100	1.28	4.16	0.12	0.33	0.46	0.39
Q-SCAD	200	100	0.31	4.57	0.41	0.57	0.26	0.10
Q-LASSO	50	100	12.58	3.41	0.00	0.27	0.71	2.15
Q-LASSO	100	100	15.09	4.60	0.00	0.65	0.46	0.67
Q-LASSO	200	100	12.56	4.84	0.00	0.84	0.30	0.24
LS-MCP	50	100	13.18	3.35	0.00	0.22	0.68	4.65
LS-MCP	100	100	22.31	4.14	0.00	0.23	0.41	3.57
LS-MCP	200	100	5.21	4.15	0.01	0.15	0.20	0.83
LS-SCAD	50	100	13.46	3.44	0.00	0.19	0.63	4.04
LS-SCAD	100	100	12.56	4.11	0.00	0.17	0.31	1.34
LS-SCAD	200	100	3.52	4.17	0.01	0.17	0.17	0.72
LS-LASSO	50	100	2.49	2.80	0.00	0.06	0.52	2.95
LS-LASSO	100	100	7.00	4.04	0.00	0.11	0.28	1.19
LS-LASSO	200	100	9.69	4.25	0.00	0.25	0.18	0.85
Q-MCP	50	300	18.94	2.96	0.00	0.11	0.68	3.26
Q-MCP	100	300	1.25	3.83	0.07	0.19	0.52	0.49
Q-MCP	200	300	0.43	4.58	0.41	0.59	0.28	0.11
Q-SCAD	50	300	12.12	2.80	0.00	0.11	0.73	2.63
Q-SCAD	100	300	1.41	3.71	0.01	0.12	0.51	0.52
Q-SCAD	200	300	0.27	4.44	0.27	0.45	0.33	0.12
Q-LASSO	50	300	19.19	3.18	0.00	0.16	0.73	2.45
Q-LASSO	100	300	20.49	4.17	0.00	0.29	0.56	1.09
Q-LASSO	200	300	18.31	4.72	0.00	0.72	0.37	0.33
LS-MCP	50	300	13.75	2.94	0.00	0.05	0.62	4.01
LS-MCP	100	300	29.59	3.99	0.00	0.20	0.47	3.64
LS-MCP	200	300	25.15	4.10	0.00	0.11	0.25	1.69
LS-SCAD	50	300	16.66	3.07	0.00	0.10	0.63	3.75
LS-SCAD	100	300	26.64	4.04	0.00	0.17	0.34	1.94
LS-SCAD	200	300	9.29	4.13	0.00	0.13	0.21	0.77
LS-LASSO	50	300	2.44	2.23	0.00	0.00	0.57	3.74
LS-LASSO	100	300	8.17	3.96	0.00	0.07	0.31	1.43
LS-LASSO	200	300	12.96	4.15	0.00	0.15	0.21	0.89
Q-MCP	50	600	19.53	2.29	0.00	0.09	0.73	3.38
Q-MCP	100	600	1.86	3.62	0.02	0.15	0.54	0.61
Q-MCP	200	600	0.27	4.51	0.36	0.52	0.30	0.11
Q-SCAD	50	600	16.61	2.19	0.00	0.07	0.79	3.72
Q-SCAD	100	600	1.00	3.48	0.00	0.12	0.59	0.63
Q-SCAD	200	600	0.35	4.43	0.28	0.45	0.35	0.14
Q-LASSO	50	600	18.01	2.46	0.00	0.08	0.86	2.96
Q-LASSO	100	600	25.36	4.08	0.00	0.31	0.53	1.08
Q-LASSO	200	600	20.91	4.71	0.00	0.72	0.41	0.41
LS-MCP	50	600	14.06	2.53	0.00	0.12	0.65	4.06
LS-MCP	100	600	29.63	3.79	0.00	0.12	0.43	2.96
LS-MCP	200	600	48.06	4.12	0.00	0.14	0.30	2.57
LS-SCAD	50	600	18.33	2.49	0.00	0.06	0.67	4.22
LS-SCAD	100	600	32.07	3.97	0.00	0.16	0.35	1.97
LS-SCAD	200	600	18.53	4.08	0.00	0.08	0.20	0.85
LS-LASSO	50	600	1.90	1.77	0.00	0.00	0.63	4.27
LS-LASSO	100	600	8.08	3.85	0.00	0.05	0.30	1.50
LS-LASSO	200	600	14.75	4.05	0.00	0.05	0.21	0.95

TABLE 1

Additional simulation results for estimating the 0.7 conditional quantile with heteroscedastic errors.

Simulation	Setting Method	n	p	FV	TV	True	P	AADE	MSE
$\epsilon \sim N(0,1)$	Q-MCP	300	100	0.46	4.00	0.77	0.01	0.15	0.04
$\epsilon \sim N(0,1)$	Q-MCP	300	300	0.17	4.00	0.86	0.00	0.16	0.02
$\epsilon \sim N(0,1)$	Q-MCP	300	600	0.17	4.00	0.87	0.00	0.15	0.03
$\epsilon \sim N(0,1)$	LS-MCP	300	100	1.41	4.00	0.70	0.03	0.13	0.03
$\epsilon \sim N(0,1)$	LS-MCP	300	300	4.50	4.00	0.67	0.02	0.13	0.11
$\epsilon \sim N(0,1)$	LS-MCP	300	600	106.34	4.00	0.00	0.16	0.21	2.11
$\epsilon \sim T_3$	Q-MCP	300	100	0.13	4.00	0.89	0.00	0.16	0.03
$\epsilon \sim T_3$	Q-MCP	300	300	0.13	4.00	0.90	0.00	0.17	0.03
$\epsilon \sim T_3$	Q-MCP	300	600	0.12	4.00	0.92	0.00	0.16	0.03
$\epsilon \sim T_3$	LS-MCP	300	100	1.14	3.99	0.60	0.02	0.20	0.06
$\epsilon \sim T_3$	LS-MCP	300	300	2.10	3.98	0.55	0.00	0.20	0.24
$\epsilon \sim T_3$	LS-MCP	300	600	3.86	3.97	0.62	0.01	0.20	1.33
Heteroscedastic $\tau = .7$	Q-MCP	300	100	0.16	4.97	0.87	0.97	0.14	0.03
Heteroscedastic $\tau = .7$	Q-MCP	300	300	0.13	4.88	0.79	0.88	0.16	0.04
Heteroscedastic $\tau = .7$	Q-MCP	300	600	0.20	4.72	0.58	0.72	0.22	0.07
Heteroscedastic $\tau = .7$	LS-MCP	300	100	2.01	4.07	0.01	0.07	0.16	0.69
Heteroscedastic $\tau = .7$	LS-MCP	300	300	4.24	4.00	0.00	0.00	0.16	0.70
Heteroscedastic $\tau = .7$	LS-MCP	300	600	11.31	4.06	0.00	0.06	0.17	0.84
Heteroscedastic $\tau = .9$	Q-MCP	300	100	0.17	5.00	0.86	1.00	0.19	0.24
Heteroscedastic $\tau = .9$	Q-MCP	300	300	0.55	4.98	0.68	1.00	0.21	0.26
Heteroscedastic $\tau = .9$	Q-MCP	300	600	0.64	4.99	0.69	1.00	0.23	0.30
Heteroscedastic $\tau = .9$	LS-MCP	300	100	2.01	4.07	0.01	0.07	0.16	4.68
Heteroscedastic $\tau = .9$	LS-MCP	300	300	4.24	4.00	0.00	0.00	0.16	4.68
Heteroscedastic $\tau = .9$	LS-MCP	300	600	11.31	4.06	0.00	0.06	0.17	4.57

TABLE 2

Simulation results using the MCP penalty function for the four simulation settings in the main paper

Method	n	p	FV	TV	True	P	AADE	MSE
Q-SCAD	300	1200	0.13	4.56	0.45	0.56	0.27	0.09
Q-SCAD	300	2400	0.13	4.42	0.34	0.43	0.33	0.12
Q-LASSO	300	1200	22.78	4.86	0.00	0.86	0.36	0.28
Q-LASSO	300	2400	26.83	4.76	0.00	0.76	0.40	0.34
LS-SCAD	300	1200	10.84	4.04	0.00	0.04	0.17	0.69
LS-SCAD	300	2400	16.77	4.04	0.00	0.04	0.16	0.67
LS-LASSO	300	1200	19.36	4.06	0.00	0.06	0.17	0.85
LS-LASSO	300	2400	20.86	4.05	0.00	0.05	0.17	0.87

TABLE 3

Extension of Table 3 from the main article with p = 1200 and 2400

τ	Method	Original NZ	Prediction Error	Randomized NZ		
0.1	Q-MCP	2.00	0.21	0.04	7.20	3.80
0.3	Q-MCP	7.00	0.21	0.04	8.84	5.32
0.5	Q-MCP	5.00	0.20	0.04	5.87	3.64
mean	LS-MCP	12.00	0.19	0.04	5.99	1.37

TABLE 4

Analysis of birth weight data using MCP penalty function

Q-SCAD .1		Q-SCAD .3		Q-SCAD .5	
Covariate	Frequency	Covariate	Frequency	Covariate	Frequency
Gestational Age	72	Gestational Age	82	Gestational Age	72
ILMN_1687073	0	ILMN_1658821	24	ILMN_1732467	56
		ILMN_1755657	23	ILMN_2334204	54
		ILMN_1804451	16	ILMN_1656361	29
		ILMN_2059464	11	ILMN_1747184	4
		ILMN_2148497	4		
		ILMN_2280960	2		

TABLE 5

Frequency of variables selected at three quantiles among 100 random partitions of the birth weight data using MCP penalty function

Method	p	FV	TV	True	L_2 error
Q-group-MCP	300	1.51	4	0.49	0.14
Q-ind-MCP	300	0.57	4	0.7	0.15
Q-group-MCP	600	0.4	4	0.67	0.1
Q-ind-MCP	600	0.79	4	0.53	0.16

TABLE 6

Simulation results with MCP penalty function corresponding to Table 7 of the main article.

Let $g_j^*(z_{ij}) = \boldsymbol{\pi}(z_{ij})'\hat{\boldsymbol{\xi}}_j$. Let $\overline{g}_j(z_{ij}) = g_j^*(z_{ij}) - E(g_j^*)$ and $\overline{g}_0 = \hat{\xi}_0 + \sum_{j=1}^d E(g_j^*)$. It is easy to see that $\hat{g}_j(z_{ij}) = \overline{g}_j(z_{ij}) - E_n\overline{g}_j(z_{ij})$, where $E_n\overline{g}_j(z_{ij}) = n^{-1}\sum_{i=1}^n \overline{g}_j(z_{ij})$; and $\hat{g}_0 = \overline{g}_0 + \sum_{j=1}^d E_n\overline{g}_j(z_{ij})$.

Note that we can express $\hat{Q}(Y_i|\mathbf{x}_i, \mathbf{z}_i) = x'_{A_i}\hat{\boldsymbol{\beta}}_1 + \overline{g}_0 + \sum_{j=1}^d \overline{g}_j(z_{ij})$; alternatively, we can express it (using the theoretically centered basis functions) as $\hat{Q}(Y_i|\mathbf{x}_i, \mathbf{z}_i) = x'_{A_i}\hat{\mathbf{c}} + \tilde{g}_0 + \sum_{j=1}^d \tilde{g}_j(z_{ij})$. Note that $E(\overline{g}_j(z_{ij})) = 0$ and $E(\tilde{g}_j(z_{ij})) = 0$, $j = 1, \ldots, d$. By the identifiability of the model (follows from the identifiability of the partially linear regression model as in Robinson (1988) and a similar argument as for Lemma 1 of Xue and Yang (2006)), we must have $\hat{\mathbf{c}}_\mathbf{1} = \hat{\boldsymbol{\beta}}_\mathbf{1}$, $\overline{g}_j = \tilde{g}_j$, $j = 0, 1, \ldots, d$. This implies $\hat{g}_j(z_{ij}) = \tilde{g}_j(z_{ij}) - n^{-1}\sum_{i=1}^n \tilde{g}_j(z_{ij})$ and $\hat{g}_0 = \tilde{g}_0 + n^{-1}\sum_{i=1}^n\sum_{j=1}^d \tilde{g}_j(z_{ij})$.

Proof of Lemma 2.

PROOF. (1) The result follows because there exist finite positive constants $b_5 < b_4$ such that $b_5 k_n^{-1} \leq E(B_j^2(z_{ik})) \leq b_4 k_n^{-1}$, for all j as shown in Corollary 1 of de Boor (1976).
(2) First, similarly as Lemma A1 of Xue and Yang (2006) there exists a constant c_j such that for any $(k_n + l)$-dimensional vector $\mathbf{a}_j$ we have $E(\mathbf{a}_j^T w(z_{ij}))^2 \geq c_j ||\mathbf{a}_j||^2 k_n^{-1}$ for all n sufficiently large. Now take any nonzero J_n-dimensional vector $\mathbf{a}$ and write is as $\mathbf{a} = (a_0, \mathbf{a}'_1, \ldots, \mathbf{a}'_d)'$ such that

$a_0 \in \mathcal{R}$ and $\mathbf{a}_j \in \mathcal{R}^{k_n+l}$, for $j = 1, \ldots, d$, then it is easy to see that $E(\mathbf{a}'\mathbf{W}(\mathbf{z}_i)\mathbf{W}(\mathbf{z}_i)^T\mathbf{a}) = E(a_0 k_n^{-1/2} + \sum_{j=1}^d \mathbf{a}_j^T w(z_{ij}))^2 \geq b_2 k_n^{-1}||\mathbf{a}||^2$ for some positive constant b_2 for all n sufficiently large, by applying the above result and Lemma 1 in Stone (1985). Hence, we obtain the lower bound. The upper bound can be derived similarly.
(3) We have

$$\lambda_{\min}(W_B^2) = \lambda_{\min}(\sum_{i=1}^n f_i(0)\mathbf{W}(\mathbf{z}_i)\mathbf{W}(\mathbf{z}_i)^T) \geq c\sum_{i=1}^n \lambda_{\min}(\mathbf{W}(\mathbf{z}_i)\mathbf{W}(\mathbf{z}_i)^T)),$$

for some positive constant c, by Condition 1 and the fact $\lambda_{\min}(A+B) \geq \lambda_{\min}(A) + \lambda_{\min}(B)$ for any two symmetric matrices A and B. From the arguments in (2), we have $E(\lambda_{\min}(W_B^2)) \geq c' n k_n^{-1}$ for some positive constant c' for all n sufficiently large. The proof finishes by noting that $||W_B^{-1}|| = \lambda_{\min}^{-1/2}(W_B^2)$.
(4) The result follows from (3) and similar argument as in the proof of Lemma 5.1 in Shi and Li (1995). (5) $\sum_{i=1}^n f_i(0)\tilde{\mathbf{x}}_i\tilde{\mathbf{W}}(\mathbf{z}_i)' = n^{-1/2}X^{*'}BWW_B^{-1} = n^{-1/2}X_A(I_n - P')BWW_B^{-1}$. The result follows because $(I_n - P')BW = BW - BW = \mathbf{0}$. □

Proof of Lemma 3.

Proof. (1) $\lambda_{\max}\left(n^{-1}X^{*'}X^*\right) = \lambda_{\max}\left(n^{-1}X_A'(I-P')(I-P)X_A\right) \leq ||I-P||^2|\lambda_{\max}\left(n^{-1}X_A'X_A\right) \leq C$, with probability one, where the last inequality follows by Condition 2 and the fact $I - P$ is a projection matrix.
(2) By the definition of X^*,

$$n^{-1/2}X^* = n^{-1/2}(X_A - PX_A) = n^{-1/2}\Delta_n + n^{-1/2}(H - PX_A).$$

Consider the following weighted least squares problem. Let $\boldsymbol{\gamma}_j^* \in \mathbb{R}^{J_n}$ be defined as $\boldsymbol{\gamma}_j^* = \underset{\boldsymbol{\gamma}\in\mathbb{R}^{J_n}}{\operatorname{argmin}} \sum_{i=1}^n f_i(0)(X_{ij} - \mathbf{W}(\mathbf{z}_i)'\boldsymbol{\gamma})^2$. Let $\hat{h}_j(\mathbf{z}_i) = \mathbf{W}(\mathbf{z}_i)'\boldsymbol{\gamma}_j^*$ and notice that $\{PX_A\}_{ij} = \hat{h}_j(\mathbf{z}_i)$. Adapting the results from Stone (1985), it follows that

$$\begin{aligned}
n^{-1}||H - PX_A||^2 &= n^{-1}\lambda_{\max}\left((H-PX_A)'(H-PX_A)\right) \\
&\leq n^{-1}\text{trace}\left[(H-PX_A)'(H-PX_A)\right] \\
&= n^{-1}\sum_{i=1}^n\sum_{j=1}^{q_n}(h_j^*(z_i) - \hat{h}_j(z_i))^2 \\
&= O_p\left(q_n n^{-2r/(2r+1)}\right) = o(1),
\end{aligned}$$

by Conditions 3-5. Observing that

$$n^{-1}X^{*\prime}B_nX^* = n^{-1}\Delta_n' B_n\Delta_n + n^{-1}\Delta_n' B_n o_p(1) + o_p(1),$$

the second conclusion follows easily. $\square$

LEMMA B.1. *Let $d_n = J_n + q_n$. If Conditions 1-5 are satisfied, then for any positive constant L,*

$$d_n^{-1} \sup_{||\boldsymbol{\theta}||\le L} \left| \sum_{i=1}^n D_i(\boldsymbol{\theta}, \sqrt{d_n}) \right| = o_p(1).$$

PROOF. Let F_{n1} denote the event $\tilde{\mathbf{s}}_{(n)} \le \alpha_1\sqrt{d_n/n}$, for some positive constant α_1, where $\tilde{\mathbf{s}}_{(n)} = \max_i ||\tilde{\mathbf{s}}_i||$. Note that $\max_i ||\tilde{\mathbf{x}}_i|| \le \alpha_2\sqrt{\frac{q_n}{n}}$, for some positive constant α_2, by Condition 2. This observation combined with Lemma 2(4) implies that $P(F_{n1}) \to 1$ as $n \to \infty$. Let F_{n2} denote the event $\max_i |u_{ni}| \le \alpha_3 k_n^{-r}$, for some positive constant α_3, then it follows from Schumaker (1981) that $P(F_{n2}) \to 1$.

To prove the lemma, it is sufficient to show that $\forall\ \epsilon > 0$,

$$P\left(d_n^{-1} \sup_{||\boldsymbol{\theta}||\le 1} \left| \sum_{i=1}^n D_i\left(\boldsymbol{\theta}, L\sqrt{d_n}\right) \right| > \epsilon, F_{n1}\cap F_{n2} \right) \to 0. \tag{B.1}$$

Define $\Theta^* \equiv \left\{\boldsymbol{\theta} \mid ||\boldsymbol{\theta}|| \le 1,\, \boldsymbol{\theta} \in \mathbb{R}^{d_n}\right\}$. We can partition Θ as a union of disjoint regions $\Theta_1, ..., \Theta_{M_n}$, such that the diameter of each region does not exceed $m_0 = \frac{\epsilon}{4\alpha_1 L\sqrt{n}}$.This covering can be constructed such that $M_n \le C\left(\frac{C\sqrt{n}}{\epsilon}\right)^{d_n+1}$, where C is a positive constant. Let $\boldsymbol{\theta}_1^*, ..., \boldsymbol{\theta}_{M_n}^*$ be arbitrary points in $\Theta_1, ..., \Theta_{M_n}$, respectively, $k = 1, \ldots, M_n$. Then

$$\begin{aligned}
& P\left(\sup_{||\boldsymbol{\theta}||\le 1} d_n^{-1} \left| \sum_{i=1}^n D_i(\boldsymbol{\theta}, L\sqrt{d_n}) \right| > \epsilon, F_{n1}\cap F_{n2} \right) \\
\le\ & \sum_{k=1}^{M_n} P\left(\sup_{\boldsymbol{\theta}\in\Theta_k} d_n^{-1} \left| \sum_{i=1}^n D_i(\boldsymbol{\theta}, L\sqrt{d_n}) \right| > \epsilon, F_{n1}\cap F_{n2} \right) \\
\le\ & \sum_{k=1}^{M_n} P\Bigg(\left| \sum_{i=1}^n D_i(\boldsymbol{\theta}_k^*, L\sqrt{d_n}) \right| + \sup_{\boldsymbol{\theta}\in\Theta_k} \left| \sum_{i=1}^n \left(D_i(\boldsymbol{\theta}, L\sqrt{d_n}) - D_i(\boldsymbol{\theta}_k^*, L\sqrt{d_n}) \right) \right| \\
& > d_n\epsilon, F_{n1}\cap F_{n2} \Bigg)
\end{aligned}$$

Let $I(\cdot)$ denote the indicator function, we will next show that

$$\sup_{\boldsymbol{\theta}\in\Theta_k}\left|d_n^{-1}\sum_{i=1}^n\left[D_i(\boldsymbol{\theta},L\sqrt{d_n})-D_i(\boldsymbol{\theta}_k^*,L\sqrt{d_n})\right]\right| I\left(F_{n1}\cap F_{n2}\right)\le \epsilon/2.$$

Write $\boldsymbol{\theta}_k^* = \left({\boldsymbol{\theta}_{k1}^*}',{\boldsymbol{\theta}_{k2}^*}'\right)'$, $k=1,\dots,M_n$. Using the triangle inequality, and the earlier derived bounds for $||\tilde{\mathbf{x}}_i||$ and $||\tilde{\mathbf{W}}(\mathbf{z}_i)||$, we have

$$\begin{aligned}
&\sup_{\boldsymbol{\theta}\in\Theta_k}\left|d_n^{-1}\sum_{i=1}^n\Big(D_i(\boldsymbol{\theta},L\sqrt{d_n})-D_i(\boldsymbol{\theta}_k^*,L\sqrt{d_n})\Big)\right| I\left(F_{n1}\cap F_{n2}\right)\\
=\;& d_n^{-1}\sup_{\boldsymbol{\theta}\in\Theta_k}\left|\sum_{i=1}^n\frac{1}{2}\left[\left|\epsilon_i-\tilde{\mathbf{x}}_i'\boldsymbol{\theta}_1L\sqrt{d_n}-\tilde{\mathbf{W}}(\mathbf{z}_i)'\boldsymbol{\theta}_2L\sqrt{d_n}-u_{ni}\right|-|\epsilon_i-u_{ni}|\right]\right.\\
&-\sum_{i=1}^n\frac{1}{2}E_s\left[\left|\epsilon_i-\tilde{\mathbf{x}}_i'\boldsymbol{\theta}_1L\sqrt{d_n}-\tilde{\mathbf{W}}(\mathbf{z}_i)'\boldsymbol{\theta}_2L\sqrt{d_n}-u_{ni}\right|-|\epsilon_i-u_{ni}|\right]\\
&+\sum_{i=1}^n L\sqrt{d_n}\Big(\tilde{\mathbf{x}}_i'\boldsymbol{\theta}_1+\tilde{\mathbf{W}}(\mathbf{z}_i)'\boldsymbol{\theta}_2\Big)\psi_\tau(\epsilon_i)\\
&-\sum_{i=1}^n\frac{1}{2}\left[\left|\epsilon_i-\tilde{\mathbf{x}}_i'\boldsymbol{\theta}_{k1}^*L\sqrt{d_n}-\tilde{\mathbf{W}}(\mathbf{z}_i)'\boldsymbol{\theta}_{k2}^*L\sqrt{d_n}-u_{ni}\right|-|\epsilon_i-u_{ni}|\right]\\
&+\sum_{i=1}^n\frac{1}{2}E_s\left[\left|\epsilon_i-\tilde{\mathbf{x}}_i'\boldsymbol{\theta}_{k1}^*L\sqrt{d_n}-\tilde{\mathbf{W}}(\mathbf{z}_i)'\boldsymbol{\theta}_{k2}^*L\sqrt{d_n}-u_{ni}\right|-|\epsilon_i-u_{ni}|\right]\\
&\left.-\sum_{i=1}^n L\sqrt{d_n}\Big(\tilde{\mathbf{x}}_i'\boldsymbol{\theta}_{k1}^*+\tilde{\mathbf{W}}(\mathbf{z}_i)'\boldsymbol{\theta}_{k2}^*\Big)\psi_\tau(\epsilon_i)\right| I\left(F_{n1}\cap F_{n2}\right)\\
\le\;& 2nLm_0d_n^{-1/2}\tilde{\mathbf{s}}_{(n)}I\left(F_{n1}\cap F_{n2}\right)\\
\le\;& 2\alpha_1nLm_0d_n^{-1/2}\sqrt{d_n/n}=2\alpha_1L\sqrt{n}m_0=\epsilon/2,
\end{aligned}$$

by the definition of m_0.

The proof of (B.1) is complete if we can verify

$$\sum_{k=1}^{M_n}P\left(\left|\sum_{i=1}^n D_i(\boldsymbol{\theta}_k^*,L\sqrt{d_n})\right|>d_n\epsilon/2,F_{n1}\cap F_{n2}\right)\to 0. \tag{B.2}$$

Applying the definition of D_i in terms of Q_i^* and the triangle inequality,

$$
\begin{aligned}
& \max_i \left| D_i(\boldsymbol{\theta}_k^*, L\sqrt{d_n}) \right| I\left(F_{n1} \cap F_{n2}\right) \\
\leq\ & \max_i \left| \left|\epsilon_i - \tilde{\mathbf{x}}_i' \boldsymbol{\theta}_{k1}^* L\sqrt{d_n} - \tilde{\mathbf{W}}(\mathbf{z}_i)' \boldsymbol{\theta}_{k2}^* L\sqrt{d_n} - u_{ni}\right| - \left|\epsilon_i - u_{ni}\right| \right| I\left(F_{n1} \cap F_{n2}\right) \\
& + \max_i \left| L\sqrt{d_n} \left(\tilde{\mathbf{x}}_i' \boldsymbol{\theta}_{k1}^* + \tilde{\mathbf{W}}(\mathbf{z}_i)' \boldsymbol{\theta}_{k2}^* \right) \psi_\tau(\epsilon_i) \right| I\left(F_{n1} \cap F_{n2}\right) \\
\leq\ & 2L\sqrt{d_n} \tilde{\mathbf{s}}_{(n)} I\left(F_{n1} \cap F_{n2}\right) \leq C d_n n^{-1/2},
\end{aligned}
$$

for some positive constant C. Define

$$
V_i(\boldsymbol{\theta}_k^*, a_n) = Q_i(a_n) - Q_i(0) + a_n(\tilde{\mathbf{x}}_i' \boldsymbol{\theta}_{k1}^* + \tilde{\mathbf{W}}(\mathbf{z}_i)' \boldsymbol{\theta}_{k2}^*) \psi_\tau(\epsilon_i).
$$

It follows from (6.2) that $D_i(\boldsymbol{\theta}_k^*, a_n) = V_i(\boldsymbol{\theta}_k^*, a_n) - E\left[V_i(\boldsymbol{\theta}_k^*, a_n) | \mathbf{x}_i,\ \mathbf{z}_i\right]$, and that

$$
\sum_{i=1}^n \operatorname{Var}\left(D_i(\boldsymbol{\theta}_k^*, a_n) I\left(F_{n1} \cap F_{n2}\right) | \mathbf{x}_i, \mathbf{z}_i\right) \leq \sum_{i=1}^n E\left[V_i^2(\boldsymbol{\theta}_k^*, a_n) I\left(F_{n1} \cap F_{n2}\right) | \mathbf{x}_i, \mathbf{z}_i\right].
$$

It follows from the Knight's identity, as presented in formula (4.3) of Koenker (2005), that

$$
\begin{aligned}
& V_i(\boldsymbol{\theta}_k^*, L\sqrt{d_n}) \\
=\ & L\sqrt{d_n} \left(\tilde{\mathbf{x}}_i' \boldsymbol{\theta}_{k1}^* + \tilde{\mathbf{W}}(\mathbf{z}_i)' \boldsymbol{\theta}_{k2}^* \right) \left[I(\epsilon_i - u_{ni} < 0) - I(\epsilon_i < 0) \right] \\
& + \int_0^{L\sqrt{d_n}(\tilde{\mathbf{x}}_i' \boldsymbol{\theta}_{k1}^* + \tilde{\mathbf{W}}(\mathbf{z}_i)' \boldsymbol{\theta}_{k2}^*)} \left[I(\epsilon_i - u_{ni} < s) - I(\epsilon_i - u_{ni} < 0) \right] ds \\
\equiv\ & V_{i1} + V_{i2}.
\end{aligned}
$$

We have

$$
\begin{aligned}
& \sum_{i=1}^n E\left[V_{i1}^2 I\left(F_{n1} \cap F_{n2}\right) | \mathbf{x}_i, \mathbf{z}_i \right] \\
=\ & \sum_{i=1}^n E\left[d_n L^2 (\tilde{\mathbf{x}}_i' \boldsymbol{\theta}_{k1}^* + \tilde{\mathbf{W}}(\mathbf{z}_i)' \boldsymbol{\theta}_{k2}^*)^2 \left| I(\epsilon_i - u_{ni} < 0) - I(\epsilon_i < 0) \right| I\left(F_{n1} \cap F_{n2}\right) \big| \mathbf{x}_i, \mathbf{z}_i \right] \\
\leq\ & 2L^2 d_n \sum_{i=1}^n E\left[\tilde{\mathbf{s}}_{(n)}^2 I\left(0 \leq |\epsilon_i| \leq |u_{ni}|\right) I\left(F_{n1} \cap F_{n2}\right) \big| \mathbf{x}_i, \mathbf{z}_i \right] \\
\leq\ & C d_n^2 n^{-1} \sum_{i=1}^n \int_{-|u_{ni}|}^{|u_{ni}|} f_i(s) ds \leq C d_n^2 k_n^{-r},
\end{aligned}
$$

for some positive constant C, where the last inequality uses Condition 1. Noting that V_{i2} is always nonnegative and that $\max_i \left|\sqrt{d_n} L\left(\tilde{\mathbf{x}}_i'\boldsymbol{\theta}_{k1}^* + \tilde{\mathbf{W}}(\mathbf{z}_i)'\boldsymbol{\theta}_{k2}^*\right)\right| \leq \alpha_1 d_n n^{-1/2}$, we have

$$
\begin{aligned}
&\sum_{i=1}^n E\left[V_{i2}^2 I\left(F_{n1}\cap F_{n2}\right)\Big|\mathbf{x}_i, \mathbf{z}_i\right] \\
\leq\ & C d_n n^{-1/2}\sum_{i=1}^n \int_0^{\sqrt{d_n}L\left(\tilde{\mathbf{x}}_i'\boldsymbol{\theta}_{k1}^*+\tilde{\mathbf{W}}(\mathbf{z}_i)'\boldsymbol{\theta}_{k2}^*\right)} \left[F_i(s+u_{ni}) - F_i(u_{ni})\right] I\left(F_{n1}\cap F_{n2}\right) ds \\
\leq\ & C d_n n^{-1/2}\sum_{i=1}^n \int_0^{\sqrt{d_n}L\left(\tilde{\mathbf{x}}_i'\boldsymbol{\theta}_{k1}^*+\tilde{\mathbf{W}}(\mathbf{z}_i)'\boldsymbol{\theta}_{k2}^*\right)} (f_i(0)+o(1))(s+O(s^2))ds \\
\leq\ & C d_n^2 n^{-1/2}\Big[\boldsymbol{\theta}_{k1}^{*\prime}(\sum_{i=1}^n f_i(0)\tilde{\mathbf{x}}_i\tilde{\mathbf{x}}_i')\boldsymbol{\theta}_{k1}^* + \boldsymbol{\theta}_{k2}^{*\prime}(\sum_{i=1}^n f_i(0)\tilde{\mathbf{W}}(\mathbf{z}_i)\tilde{\mathbf{W}}(\mathbf{z}_i)')\boldsymbol{\theta}_{k2}^*\Big](1+o(1)) \\
\leq\ & C d_n^2 n^{-1/2}\left[||\boldsymbol{\theta}_{k1}^*||^2 \lambda_{\max}(n^{-1}X^{*'}B_n X^*) + ||\boldsymbol{\theta}_{k2}^*||^2\right](1+o(1)) \\
\leq\ & C d_n^2 n^{-1/2}(1+o(1)),
\end{aligned}
$$

for some positive constant C. The second to last inequality follows because $\sum_{i=1}^n f_i(0)\tilde{\mathbf{W}}(\mathbf{z}_i)\tilde{\mathbf{W}}(\mathbf{z}_i)' = W_B^{-1}W^T BWW_B^{-1} = W_B^{-1}W_B^2 W_B^{-1} = I$. The last inequality follows from Lemma 3 and Condition 1. Therefore

$$
\sum_{i=1}^n \mathrm{Var}\Big(D_i(\boldsymbol{\theta}) I\left(F_{n1}\cap F_{n2}\right)\Big|\mathbf{x}_i, \mathbf{z}_i\Big) \leq C d_n^2 k_n^{-r},
$$

for some positive constant C and all n sufficiently large. By Bernstein's inequality, for all n sufficiently large,

$$
\begin{aligned}
&\sum_{k=1}^{M_n} P\left(\left|\sum_{i=1}^n D_i(\boldsymbol{\theta}_k^*, L\sqrt{d_n/n})\right| > d_n\epsilon/2, F_{n1}\cap F_{n2}\Big|\mathbf{x}_i, \mathbf{z}_i\right) \\
\leq\ & 2\sum_{k=1}^{M_n}\exp\left(\frac{-d_n^2\epsilon^2/4}{C d_n^2 k_n^{-r} + C\epsilon d_n n^{-1/2}}\right) \\
\leq\ & 2\sum_{k=1}^{M_n}\exp\left(-Ck_n^r\right) = 2M_n \exp\left(-Ck_n^r\right) \\
\leq\ & C\left(\frac{C\sqrt{n}}{\epsilon}\right)^{d_n+1}\exp\left(-Ck_n^r\right) \\
=\ & C\exp\left((d_n+1)\log(C\sqrt{n}/\epsilon) - Ck_n^r\right) \\
\leq\ & C\exp\left(C(d_n+1)\log(n) - Ck_n^r\right),
\end{aligned}
$$

which converges to zero as $n \to \infty$ by Conditions 3-5. Note that the upper bound does not depend on $\{\mathbf{x}_i, \mathbf{z}_i\}$, so the above bound also holds unconditionally. This implies (B.2). Hence, the proof is complete. $\square$

Proof of Lemma 4.

PROOF. We will first prove that $\forall\, \eta > 0$, there exists an $L > 0$ such that

$$P\left(\inf_{||\boldsymbol{\theta}||=L} d_n^{-1} \sum_{i=1}^{n} (Q_i(\sqrt{d_n}) - Q_i(0)) > 0\right) \geq 1 - \eta. \tag{B.3}$$

Note that

$$\begin{aligned} d_n^{-1} \sum_{i=1}^{n} (Q_i(\sqrt{d_n}) - Q_i(0)) &= d_n^{-1} \sum_{i=1}^{n} D_i(\boldsymbol{\theta}, \sqrt{d_n}) + d_n^{-1} \sum_{i=1}^{n} E_s[Q_i(\sqrt{d_n}) - Q_i(0)] \\ &\quad - d_n^{-1/2} \sum_{i=1}^{n} (\tilde{\mathbf{x}}_i' \boldsymbol{\theta}_1 + \tilde{\mathbf{W}}(\mathbf{z}_i)' \boldsymbol{\theta}_2) \psi_\tau(\epsilon_i) \\ &= G_{n1} + G_{n2} + G_{n3}, \end{aligned}$$

where the definition of G_{ni}, $i = 1, 2, 3$, is clear from the context. First, we note that $\sup_{||\boldsymbol{\theta}|| \leq L} |G_{n1}| = o_p(1)$ by Lemma B.1. We next evaluate G_{n3}. Noting that $E(G_{n3}) = 0$ and that by Condition 1 there exists some constant C such that $\boldsymbol{\theta}_2' \sum_i \tilde{\mathbf{W}}(\mathbf{z}_i) \tilde{\mathbf{W}}(\mathbf{z}_i)' \boldsymbol{\theta}_2 \leq C \boldsymbol{\theta}_2' \sum_i f_i(0) \tilde{\mathbf{W}}(\mathbf{z}_i) \tilde{\mathbf{W}}(\mathbf{z}_i)' \boldsymbol{\theta}_2 = C||\boldsymbol{\theta}_2||^2$. Then by Lemma 3 we have

$$E(G_{n3}^2) \leq C d_n^{-1} \mathrm{E}\left[n^{-1} \boldsymbol{\theta}_1' X^{*'} X^* \boldsymbol{\theta}_1 + ||\boldsymbol{\theta}_2||^2\right] = O\left(d_n^{-1} ||\boldsymbol{\theta}||^2\right).$$

Therefore $G_{n3} = O_p\left(d_n^{-1/2} ||\boldsymbol{\theta}||\right)$. Next, we analyze G_{n2}. Applying Knight's identity twice, we have

$$\begin{aligned}
G_{n2} &= d_n^{-1}\sum_{i=1}^n E\left[\int_{u_{ni}}^{\sqrt{d_n}(\tilde{\mathbf{x}}_i'\boldsymbol{\theta}_1+\tilde{\mathbf{W}}(\mathbf{z}_i)'\boldsymbol{\theta}_2)+u_{ni}}[I(\epsilon_i<s)-I(\epsilon_i<0)]ds\Big|\mathbf{x}_i,\mathbf{z}_i\right]\\
&= d_n^{-1}\sum_{i=1}^n \int_{u_{ni}}^{\sqrt{d_n}(\tilde{\mathbf{x}}_i'\boldsymbol{\theta}_1+\tilde{\mathbf{W}}(\mathbf{z}_i)'\boldsymbol{\theta}_2)+u_{ni}} f_i(0)s\,ds(1+o(1))\\
&= d_n^{-1}\sum_{i=1}^n f_i(0)\left[\frac{1}{2}d_n\left(\tilde{\mathbf{x}}_i'\boldsymbol{\theta}_1+\tilde{\mathbf{W}}(\mathbf{z}_i)'\boldsymbol{\theta}_2\right)^2+u_{ni}\sqrt{d_n}\left(\tilde{\mathbf{x}}_i'\boldsymbol{\theta}_1+\tilde{\mathbf{W}}(\mathbf{z}_i)'\boldsymbol{\theta}_2\right)\right]\\
&\quad\times(1+o(1))\\
&= C\boldsymbol{\theta}_1'\big(n^{-1}\sum_{i=1}^n f_i(0)\mathbf{x}_i^*\mathbf{x}_i^{*\prime}\big)\boldsymbol{\theta}_1\times(1+o(1))\\
&\quad +C\boldsymbol{\theta}_2'\left(\sum_{i=1}^n f_i(0)\tilde{\mathbf{W}}(\mathbf{z}_i)\tilde{\mathbf{W}}(\mathbf{z}_i)'\right)\boldsymbol{\theta}_2(1+o(1))\\
&\quad +d_n^{-1/2}\sum_{i=1}^n f_i(0)u_{ni}\left(\tilde{\mathbf{x}}_i'\boldsymbol{\theta}_1+\tilde{\mathbf{W}}(\mathbf{z}_i)'\boldsymbol{\theta}_2\right)\\
&= Cn^{-1}\boldsymbol{\theta}_1'X^{*\prime}B_nX^*\boldsymbol{\theta}_1(1+o(1))+C||\boldsymbol{\theta}_2||^2(1+o(1))\\
&\quad +d_n^{-1/2}\sum_{i=1}^n f_i(0)u_{ni}\left(\tilde{\mathbf{x}}_i'\boldsymbol{\theta}_1+\tilde{\mathbf{W}}(\mathbf{z}_i)'\boldsymbol{\theta}_2\right),
\end{aligned}$$

where the second to last equality follows from Lemma 2(5).We note that by Lemma 3, $n^{-1}\boldsymbol{\theta}_1'X^{*\prime}B_nX^*\boldsymbol{\theta}_1=\boldsymbol{\theta}_1'K_n\boldsymbol{\theta}_1+o_p(||\boldsymbol{\theta}_1||^2)$. Hence, by Condition 2, there exists a finite constant $c>0$, such that $Cn^{-1}\boldsymbol{\theta}_1'X^{*\prime}B_nX^*\boldsymbol{\theta}_1(1+o(1))+C||\boldsymbol{\theta}_2||^2(1+o(1))\geq c||\boldsymbol{\theta}||^2$ with probability approaching one. Let $U_n=(u_{n1},...,u_{nn})'$. By Schumaker (1981), $||U_n||=O(\sqrt{n}k_n^{-r})$. By Condition 2 and Cauchy-Schwarz inequality, we have

$$\begin{aligned}
d_n^{-1/2}\sum_{i=1}^n f_i(0)u_{ni}\tilde{\mathbf{x}}_i'\boldsymbol{\theta}_1 &= d_n^{-1/2}n^{-1/2}\boldsymbol{\theta}_1'X^{*\prime}B_nU_n\\
&\leq d_n^{-1/2}||n^{-1/2}\boldsymbol{\theta}_1'X^{*\prime}||\cdot||B_nU_n||\\
&= O_p(d_n^{-1/2}\sqrt{n}k_n^{-r})||\boldsymbol{\theta}||=O_p(||\boldsymbol{\theta}||).
\end{aligned}$$

Similarly,

$$\begin{aligned}
d_n^{-1/2}\sum_{i=1}^n f_i(0)u_{ni}\tilde{\mathbf{W}}(\mathbf{z}_i)'\boldsymbol{\theta}_2 &= d_n^{-1/2}\boldsymbol{\theta}_2'W_B^{-1}W'B_nU_n\\
&\leq d_n^{-1/2}||\boldsymbol{\theta}_2'W_B^{-1}W'B_n^{1/2}||\cdot||B_n^{1/2}U_n||=O_p(||\boldsymbol{\theta}||).
\end{aligned}$$

Hence, for L sufficiently large, the quadratic term will dominate and $d_n^{-1}\sum_{i=1}^n(Q_i(\sqrt{d_n})-Q_i(0))$ has asymptotically a lower bound cL^2. This proves (B.3) and by convexity implies $||\hat{\boldsymbol\theta}|| = O_p(\sqrt{d_n})$, where $\hat{\boldsymbol\theta} = (\hat{\boldsymbol\theta}_1', \hat{\boldsymbol\theta}_2')'$. From the definition of $\hat{\boldsymbol\theta}$, it follows that $||W_B(\hat\gamma-\gamma_0)|| = O_p\left(\sqrt{d_n}\right)$. Using these facts and Condition 4,

$$\begin{aligned} n^{-1}\sum_{i=1}^n f_i(0)\left(\tilde g(\mathbf{z}_i)-g_0(\mathbf{z}_i)\right)^2 &= n^{-1}\sum_{i=1}^n f_i(0)\left(\mathbf{W}(\mathbf{z}_i)'(\hat\gamma-\gamma_0)-u_{ni}\right)^2 \\ &\le n^{-1}C\left(\hat\gamma-\gamma_0\right)'W_B^2\left(\hat\gamma-\gamma_0\right)+O_p\left(k_n^{-2r}\right) \\ &= O_p\left(n^{-1}d_n\right). \end{aligned}$$

By Condition 1, $n^{-1}\sum_{i=1}^n\left(\tilde g(\mathbf{z}_i)-g_0(\mathbf{z}_i)\right)^2 = O_p\left(n^{-1}d_n\right).$ □

LEMMA B.2. *Assume Conditions 1-6 hold and* $\log(p_n) = o\left(n\lambda^2\right)$ *and* $n\lambda^2\to\infty$, *then*

$$P\left(\max_{q_n+1\le j\le p_n}\frac{1}{n}\left|\sum_{i=1}^n x_{ij}\left[I(Y_i-\mathbf{x}_{A_i}{}'\boldsymbol\beta_{01}-g_0(\mathbf{z}_i)\le 0)-\tau\right]\right|>\lambda/2\right)\to 0.$$

PROOF. Proof follows similarly as that for Lemma 4.2 of Wang, Wu and Li (2012) by applying Hoeffding's inequality. □

Lemmas B.3 and B.4 below will be used to show that $||\hat{\boldsymbol\theta}_1-\tilde{\boldsymbol\theta}_1|| = o_p(1)$. Define

$$\tilde Q_i(\boldsymbol\theta_1,\tilde{\boldsymbol\theta}_1,\boldsymbol\theta_2) = \rho_\tau(\epsilon_i-\tilde{\mathbf{x}}_i'\boldsymbol\theta_1-\tilde{\mathbf{W}}(\mathbf{z}_i)'\boldsymbol\theta_2-u_{ni})-\rho_\tau(\epsilon_i-\tilde{\mathbf{x}}_i'\tilde{\boldsymbol\theta}_1-\tilde{\mathbf{W}}(\mathbf{z}_i)'\boldsymbol\theta_2-u_{ni}).$$

LEMMA B.3. *If Conditions 1-5 hold, then for any finite positive constants* M *and* C,

$$\begin{aligned} &\sup_{||\boldsymbol\theta_1-\tilde{\boldsymbol\theta}_1||\le M,\,||\boldsymbol\theta_2||\le C\sqrt{d_n}}\left|\sum_{i=1}^n E_s\left[\tilde Q_i(\boldsymbol\theta_1,\tilde{\boldsymbol\theta}_1,\boldsymbol\theta_2)\right]-\frac{1}{2}\left[\boldsymbol\theta_1'K_n\boldsymbol\theta_1-\tilde{\boldsymbol\theta}_1'K_n\tilde{\boldsymbol\theta}_1\right](1+o(1))\right| \\ &= o_p(1). \end{aligned}$$

PROOF. Applying Knight's identity, we have

$$\begin{aligned}
&\sum_{i=1}^{n} E_s\left[\tilde{Q}_i(\boldsymbol{\theta}_1, \tilde{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_2)\right] \\
= &\sum_{i=1}^{n} \int_{\tilde{\mathbf{x}}_i'\tilde{\boldsymbol{\theta}}_1+\tilde{\mathbf{W}}(\mathbf{z}_i)'\boldsymbol{\theta}_2+u_{ni}}^{\tilde{\mathbf{x}}_i'\boldsymbol{\theta}_1+\tilde{\mathbf{W}}(\mathbf{z}_i)'\boldsymbol{\theta}_2+u_{ni}} (F_i(s)-F_i(0))ds \\
= &\frac{1}{2}\sum_{i=1}^{n} f_i(0)(1+o(1))\left[\left(\tilde{\mathbf{x}}_i'\boldsymbol{\theta}_1+\tilde{\mathbf{W}}(\mathbf{z}_i)'\boldsymbol{\theta}_2+u_{ni}\right)^2-\left(\tilde{\mathbf{x}}_i'\tilde{\boldsymbol{\theta}}_1+\tilde{\mathbf{W}}(\mathbf{z}_i)'\boldsymbol{\theta}_2+u_{ni}\right)^2\right] \\
= &\frac{1}{2}\sum_{i=1}^{n} f_i(0)\left[\left(\tilde{\mathbf{x}}_i'\boldsymbol{\theta}_1\right)^2-\left(\tilde{\mathbf{x}}_i'\tilde{\boldsymbol{\theta}}_1\right)^2+2\left(\tilde{\mathbf{W}}(\mathbf{z}_i)'\boldsymbol{\theta}_2+u_{ni}\right)\left(\tilde{\mathbf{x}}_i'\boldsymbol{\theta}_1-\tilde{\mathbf{x}}_i'\tilde{\boldsymbol{\theta}}_1\right)\right](1+o(1)) \\
= &\frac{1}{2}\left[\boldsymbol{\theta}_1'K_n\boldsymbol{\theta}_1-\tilde{\boldsymbol{\theta}}_1'K_n\tilde{\boldsymbol{\theta}}_1\right](1+o(1))+\frac{1}{2}(\boldsymbol{\theta}_1-\tilde{\boldsymbol{\theta}}_1)'n^{-1/2}\sum_{i=1}^{n} f_i(0)\mathbf{x}_i^*u_{ni},
\end{aligned}$$

where the last step applies Lemmas 2(5) and 3. By Lemma 3

$$\frac{1}{2}(\boldsymbol{\theta}_1-\tilde{\boldsymbol{\theta}}_1)'n^{-1/2}\sum_{i=1}^{n} f_i(0)\mathbf{x}_i^*u_{ni} = \frac{1}{2}(\boldsymbol{\theta}_1-\tilde{\boldsymbol{\theta}}_1)'n^{-1/2}\sum_{i=1}^{n} f_i(0)\boldsymbol{\delta}_i u_{ni}(1+o_p(1)).$$

Note that $E[\boldsymbol{\delta}_i]=0$ and $\sup_i |u_{ni}| = O(k_n^{-r})$, we have $\sup_{||\boldsymbol{\theta}_1-\tilde{\boldsymbol{\theta}}_1||\le M,\, ||\boldsymbol{\theta}_2||\le C\sqrt{d_n}} |\frac{1}{2}(\boldsymbol{\theta}_1-\tilde{\boldsymbol{\theta}}_1)'n^{-1/2}\sum_{i=1}^{n} f_i(0)\mathbf{x}_i^*u_{ni}| = o_p(1)$. This completes the proof. □

LEMMA B.4. *Let* $A_i(\boldsymbol{\theta}_1, \tilde{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_2) = \tilde{Q}_i(\boldsymbol{\theta}_1, \tilde{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_2) - E_s\left[\tilde{Q}_i(\boldsymbol{\theta}_1, \tilde{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_2)\right] + \tilde{\mathbf{x}}_i'\left(\boldsymbol{\theta}_1-\tilde{\boldsymbol{\theta}}_1\right)\psi_\tau(\epsilon_i)$. *If Conditions 1-5 hold, then for any given positive constants* M *and* C,

$$\sup_{||\boldsymbol{\theta}_1-\tilde{\boldsymbol{\theta}}_1||\le M,\, ||\boldsymbol{\theta}_2||\le C\sqrt{d_n}}\left|\sum_{i=1}^{n} A_i(\boldsymbol{\theta}_1, \tilde{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_2)\right| = o_p(1).$$

PROOF. Let F_{n1} and F_{n2} be the events defined in Lemma B.1. Then proof will be complete if we can verify $\forall\epsilon>0$,

$$P\left(\sup_{||\boldsymbol{\theta}_1-\tilde{\boldsymbol{\theta}}_1||\le M,\, ||\boldsymbol{\theta}_2||\le C\sqrt{d_n}}\left|\sum_{i=1}^{n} A_i(\boldsymbol{\theta}_1, \tilde{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_2)\right| > \epsilon,\, F_{n1}\cap F_{n2}\right)\to 0 \tag{B.4}$$

Similarly as in the proof of Lemma B.1 let $\Theta_1 = \left\{\boldsymbol{\theta}_1 :\ ||\boldsymbol{\theta}-\tilde{\boldsymbol{\theta}}_1|| \le M,\, \boldsymbol{\theta}_1\in\mathbb{R}^{q_n}\right\}$ and $\Theta_2 = \left\{\boldsymbol{\theta}_2 :\ ||\boldsymbol{\theta}_2|| \le C\sqrt{d_n},\, \boldsymbol{\theta}_2\in\mathbb{R}^{J_n}\right\}$. We can partition Θ_1 (similarly

Θ_2), into disjoint regions $\Theta_{11},...,\Theta_{1K_n}$ ($\Theta_{21},...,\Theta_{2L_n}$) such that the diameter of each region does not exceed $m_0^* = \frac{\epsilon}{10\alpha_1\sqrt{nd_n}}$, where α_1 is the constant defined in the proof of Lemma B.1.These partitions can be constructed such that $K_n \le C\left(\frac{C\sqrt{nd_n}}{2\epsilon}\right)^{q_n}$ and $L_n \le C\left(\frac{C\sqrt{nd_n}}{2\epsilon}\right)^{dJ_n+1}$. Let $\boldsymbol{\theta}_{11},...,\boldsymbol{\theta}_{1K_n}$ be arbitrary points in $\Theta_{11},...,\Theta_{1K_n}$, respectively; similarly, let $\boldsymbol{\theta}_{21},...,\boldsymbol{\theta}_{2L_n}$ be arbitrary points in $\Theta_{21},...,\Theta_{2L_n}$, respectively.

Then (B.4) is bounded by

$$\sum_{l=1}^{L_n}\sum_{k=1}^{K_n} P\Bigg(\big|\sum_{i=1}^n A_i(\boldsymbol{\theta}_{1k},\tilde{\boldsymbol{\theta}}_1,\boldsymbol{\theta}_{2l})\big| + \sup_{\boldsymbol{\theta}_1\in\Theta_{1k},\,\boldsymbol{\theta}_2\in\Theta_{2l}} \left|\sum_{i=1}^n \left(A_i(\boldsymbol{\theta}_1,\tilde{\boldsymbol{\theta}}_1,\boldsymbol{\theta}_2) - A_i(\boldsymbol{\theta}_{1k},\tilde{\boldsymbol{\theta}}_1,\boldsymbol{\theta}_{2l})\right)\right| > \epsilon,\, F_{n1}\cap F_{n2}\Bigg).$$

Note that

$$\begin{aligned}
& \tilde{Q}_i(\boldsymbol{\theta}_1,\tilde{\boldsymbol{\theta}}_1,\boldsymbol{\theta}_2) - \tilde{Q}_i(\boldsymbol{\theta}_{1k},\tilde{\boldsymbol{\theta}}_1,\boldsymbol{\theta}_{2l}) \\
= \;& \frac{1}{2}\Big[\Big|\epsilon_i - \tilde{\mathbf{x}}_i'\boldsymbol{\theta}_1 - \tilde{\mathbf{W}}(\mathbf{z}_i)'\boldsymbol{\theta}_2 - u_{ni}\Big| - \Big|\epsilon_i - \tilde{\mathbf{x}}_i'\boldsymbol{\theta}_{1k} - \tilde{\mathbf{W}}(\mathbf{z}_i)'\boldsymbol{\theta}_{2l} - u_{ni}\Big|\Big] \\
- \;& \frac{1}{2}\Big[\Big|\epsilon_i - \tilde{\mathbf{x}}_i'\tilde{\boldsymbol{\theta}}_1 - \tilde{\mathbf{W}}(\mathbf{z}_i)'\boldsymbol{\theta}_2 - u_{ni}\Big| - \Big|\epsilon_i - \tilde{\mathbf{x}}_i'\tilde{\boldsymbol{\theta}}_1 - \tilde{\mathbf{W}}(\mathbf{z}_i)'\boldsymbol{\theta}_{2l} - u_{ni}\Big|\Big] \\
+ \;& (\tau - 1/2)(\tilde{\mathbf{x}}_i'(\boldsymbol{\theta}_{1k} - \boldsymbol{\theta}_1)) \\
\le \;& 2\tilde{\mathbf{s}}_{(n)} \sup_{\boldsymbol{\theta}_1\in\Theta_{1k},\,\boldsymbol{\theta}_2\in\Theta_{2l}} \Big[||\boldsymbol{\theta}_1 - \boldsymbol{\theta}_{1k}|| + ||\boldsymbol{\theta}_2 - \boldsymbol{\theta}_{2l}||\Big],
\end{aligned}$$

where $\tilde{\mathbf{s}}_{(n)}$ is defined in the proof of Lemma B.1.

Using the above bound, the definition of $A_i(\boldsymbol{\theta}_1,\tilde{\boldsymbol{\theta}}_1,\boldsymbol{\theta}_2)$ and m_0^*, we have

$$\begin{aligned}
& \sup_{\boldsymbol{\theta}_1\in\Theta_{1k}\,\boldsymbol{\theta}_2\in\Theta_{2l}} \sum_{i=1}^n \Big|A_i(\boldsymbol{\theta}_1,\tilde{\boldsymbol{\theta}}_1,\boldsymbol{\theta}_2) - A_i(\bar{\boldsymbol{\theta}}_{l_1},\tilde{\boldsymbol{\theta}}_1,\bar{\boldsymbol{\theta}}_{l_2})\Big| I(F_{n1}\cap F_{n2}) \\
\le \;& 5n\max_i||\tilde{\mathbf{s}}_i|| \sup_{\boldsymbol{\theta}_1\in\Theta_{1k}\,\boldsymbol{\theta}_2\in\Theta_{2l}} \Big[||\boldsymbol{\theta}_1 - \boldsymbol{\theta}_{1k}|| + ||\boldsymbol{\theta}_2 - \boldsymbol{\theta}_{2l}||\Big] I(F_{n1}\cap F_{n2}) \\
\le \;& 5\alpha_1 m_o^*\sqrt{nd_n} = \epsilon/2.
\end{aligned}$$

Hence, the proof will be complete if we can show

$$\sum_{l=1}^{L_n}\sum_{k=1}^{K_n} P\left(\left|\sum_{i=1}^n A_i(\boldsymbol{\theta}_{1k},\tilde{\boldsymbol{\theta}}_1,\boldsymbol{\theta}_{2l})\right| > \epsilon/2, F_{n1}\cap F_{n2}\right) \to 0. \tag{B.5}$$

To prove this, we apply Bernstein's inequality. We need to evaluate the variance and maximum of $A_i(\boldsymbol{\theta}_{1k}, \tilde{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_{2l})$. Using an equality similar to (6.3) from the appendix of the main paper, we have

$$\begin{aligned}
& \max_{i,k,l} \left| A_i(\boldsymbol{\theta}_{1k}, \tilde{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_{2l}) \right| I(F_{n1} \cap F_{n2}) \\
= \; & \max_{i,k,l} \Bigg[\frac{1}{2} \Big[|\epsilon_i - \tilde{\mathbf{x}}_i' \boldsymbol{\theta}_{1k} - \tilde{\mathbf{W}}(\mathbf{z}_i)' \boldsymbol{\theta}_{2l} - u_{ni}| - |\epsilon_i - \tilde{\mathbf{x}}_i' \tilde{\boldsymbol{\theta}}_1 - \tilde{\mathbf{W}}(\mathbf{z}_i)' \boldsymbol{\theta}_{2l} - u_{ni}| \\
& + (\tau - 1/2) \left(\tilde{\mathbf{x}}_i' (\tilde{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_{1k}) \right) \Big] \\
- \; & \frac{1}{2} E_s \Big[|\epsilon_i - \tilde{\mathbf{x}}_i' \boldsymbol{\theta}_{1k} - \tilde{\mathbf{W}}(\mathbf{z}_i)' \boldsymbol{\theta}_{2l} - u_{ni}| - |\epsilon_i - \tilde{\mathbf{x}}_i' \tilde{\boldsymbol{\theta}}_1 - \tilde{\mathbf{W}}(\mathbf{z}_i)' \boldsymbol{\theta}_{2l} - u_{ni}| \\
& - (\tau - 1/2) \left(\tilde{\mathbf{x}}_i' (\tilde{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_{1k}) \right) \Big] + \tilde{\mathbf{x}}_i' (\boldsymbol{\theta}_{1k} - \tilde{\boldsymbol{\theta}}_1) \psi_\tau(\epsilon_i) \Bigg] I(F_{n1} \cap F_{n2}) \\
\leq \; & 3 \max_i ||\tilde{\mathbf{x}}_i|| \max_k ||\boldsymbol{\theta}_{1k} - \tilde{\boldsymbol{\theta}}_1|| I(F_{n1} \cap F_{n2}) \leq C \sqrt{q_n/n}.
\end{aligned}$$

Applying Knight's identity,

$$\begin{aligned}
& \tilde{Q}_i(\boldsymbol{\theta}_{1k}, \tilde{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_{2l}) + (\boldsymbol{\theta}_{1k} - \tilde{\boldsymbol{\theta}}_1)' \tilde{\mathbf{x}}_i \psi_\tau(\epsilon_i) \\
= \; & \rho_\tau(\epsilon_i - \tilde{\mathbf{x}}_i' \boldsymbol{\theta}_{1k} - \tilde{\mathbf{W}}(\mathbf{z}_i)' \boldsymbol{\theta}_{2l} - u_{ni}) - \rho_\tau(\epsilon_i - \tilde{\mathbf{x}}_i' \tilde{\boldsymbol{\theta}}_1 - \tilde{\mathbf{W}}(\mathbf{z}_i)' \boldsymbol{\theta}_{2l} - u_{ni}) + (\boldsymbol{\theta}_{1k} - \tilde{\boldsymbol{\theta}}_1)' \tilde{\mathbf{x}}_i \psi_\tau(\epsilon_i) \\
= \; & -(\tilde{\mathbf{x}}_i' \boldsymbol{\theta}_{1k} + \tilde{\mathbf{W}}(\mathbf{z}_i)' \boldsymbol{\theta}_{2l} + u_{ni}) \psi_\tau(\epsilon_i) + \int_0^{\tilde{\mathbf{x}}_i' \boldsymbol{\theta}_{1k} + \tilde{\mathbf{W}}(\mathbf{z}_i)' \boldsymbol{\theta}_{2l} + u_{ni}} [I(\epsilon_i < t) - I(\epsilon_i < 0)] dt \\
& + (\tilde{\mathbf{x}}_i' \tilde{\boldsymbol{\theta}}_1 + \tilde{\mathbf{W}}(\mathbf{z}_i)' \boldsymbol{\theta}_{2l} + u_{ni}) \psi_\tau(\epsilon_i) - \int_0^{\tilde{\mathbf{x}}_i' \tilde{\boldsymbol{\theta}}_1 + \tilde{\mathbf{W}}(\mathbf{z}_i)' \boldsymbol{\theta}_{2l} + u_{ni}} [I(\epsilon_i < t) - I(\epsilon_i < 0)] dt \\
& + (\boldsymbol{\theta}_{1k} - \tilde{\boldsymbol{\theta}}_1)' \tilde{\mathbf{x}}_i \psi_\tau(\epsilon_i) \\
= \; & \int_{\tilde{\mathbf{x}}_i' \tilde{\boldsymbol{\theta}}_1 + \tilde{\mathbf{W}}(\mathbf{z}_i)' \boldsymbol{\theta}_{2l} + u_{ni}}^{\tilde{\mathbf{x}}_i' \boldsymbol{\theta}_{1k} + \tilde{\mathbf{W}}(\mathbf{z}_i)' \boldsymbol{\theta}_{2l} + u_{ni}} [I(\epsilon_i < t) - I(\epsilon_i < 0)] dt.
\end{aligned}$$

Hence, $A_i(\boldsymbol{\theta}_{1k}, \tilde{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_{2l})) = \int_{\tilde{\mathbf{x}}_i' \tilde{\boldsymbol{\theta}}_1 + \tilde{\mathbf{W}}(\mathbf{z}_i)' \boldsymbol{\theta}_{2l} + u_{ni}}^{\tilde{\mathbf{x}}_i' \boldsymbol{\theta}_{1k} + \tilde{\mathbf{W}}(\mathbf{z}_i)' \boldsymbol{\theta}_{2l} + u_{ni}} [I(\epsilon_i < t) - I(\epsilon_i < 0)] dt - E_s[\int_{\tilde{\mathbf{x}}_i' \tilde{\boldsymbol{\theta}}_1 + \tilde{\mathbf{W}}(\mathbf{z}_i)' \boldsymbol{\theta}_{2l} + u_{ni}}^{\tilde{\mathbf{x}}_i' \boldsymbol{\theta}_{1k} + \tilde{\mathbf{W}}(\mathbf{z}_i)' \boldsymbol{\theta}_{2l} + u_{ni}} [I(\epsilon_i < t) - I(\epsilon_i < 0)] dt]$.

We have

$$
\begin{aligned}
&\sum_{i=1}^{n} \mathrm{Var}(A_i(\boldsymbol{\theta}_{1k}, \tilde{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_{2l}) I(F_{n1} \cap F_{n2}) | \mathbf{x}_i, \mathbf{z}_i) \\
\leq\ & E\Big[\Big(\int_{\tilde{\mathbf{x}}_i'\tilde{\boldsymbol{\theta}}_1 + \tilde{\mathbf{W}}(\mathbf{z}_i)'\boldsymbol{\theta}_{2l} + u_{ni}}^{\tilde{\mathbf{x}}_i'\boldsymbol{\theta}_{1k} + \tilde{\mathbf{W}}(\mathbf{z}_i)'\boldsymbol{\theta}_{2l} + u_{ni}} [I(\epsilon_i < t) - I(\epsilon_i < 0)] dt \Big)^2 \Big| \mathbf{x}_i, \mathbf{z}_i \Big] \\
\leq\ & \max_{i,k} \Big| \tilde{\mathbf{x}}_i'(\boldsymbol{\theta}_{1k} - \tilde{\boldsymbol{\theta}}_1) \Big| \sum_{i=1}^{n} \int_{\tilde{\mathbf{x}}_i'\tilde{\boldsymbol{\theta}}_1 + \tilde{\mathbf{W}}(\mathbf{z}_i)'\boldsymbol{\theta}_{2l} + u_{ni}}^{\tilde{\mathbf{x}}_i'\boldsymbol{\theta}_{1k} + \tilde{\mathbf{W}}(\mathbf{z}_i)'\boldsymbol{\theta}_{2l} + u_{ni}} [F_i(t) - F_i(0)]\, dt I(F_{n1} \cap F_{n2}) \\
\leq\ & C\sqrt{q_n} n^{-1/2} \sum_{i=1}^{n} \Big[(\tilde{\mathbf{x}}_i'\boldsymbol{\theta}_{1k} + \tilde{\mathbf{W}}(\mathbf{z}_i)'\boldsymbol{\theta}_{2l} + u_{ni})^2 - (\tilde{\mathbf{x}}_i'\tilde{\boldsymbol{\theta}}_1 + \tilde{\mathbf{W}}(\mathbf{z}_i)'\boldsymbol{\theta}_{2l} + u_{ni})^2 \Big] \\
& \times (1 + o(1)) I(F_{n1} \cap F_{n2}) \\
\leq\ & C\sqrt{q_n} n^{-1/2} \sum_{i=1}^{n} \Big[\left(\tilde{\mathbf{x}}_i'\boldsymbol{\theta}_{1k}\right)^2 - \left(\tilde{\mathbf{x}}_i'\tilde{\boldsymbol{\theta}}_1\right)^2 + 2\left(\tilde{\mathbf{W}}(\mathbf{z}_i)'\boldsymbol{\theta}_{2l} + u_{ni}\right)\left(\tilde{\mathbf{x}}_i'\boldsymbol{\theta}_{1k} - \tilde{\mathbf{x}}_i'\tilde{\boldsymbol{\theta}}_1\right) \Big] \\
& \times (1 + o(1)) I(F_{n1} \cap F_{n2}) = C\sqrt{q_n} n^{-1/2}(1 + o(1)),
\end{aligned}
$$

following the argument similarly as in the proof of Lemma B.3. Therefore

$$
\sum_{i=1}^{n} \mathrm{Var}\big(A_i(\boldsymbol{\theta}_{1k}, \tilde{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_{2l}) I(F_{n1} \cap F_{n2}) | \mathbf{x}_i, \mathbf{z}_i\big) \leq C\sqrt{q_n} n^{-1/2},
$$

for all n sufficiently large. By Bernstein's inequality and Conditions 4 and 5

$$
\begin{aligned}
&\sum_{l=1}^{L_n} \sum_{k=1}^{K_n} P\left(\left| \sum_{i=1}^{n} A_i(\boldsymbol{\theta}_{1k}, \tilde{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_{2l}) \right| > \epsilon/2, F_{n1} \cap F_{n2} \Big| \mathbf{x}_i, \mathbf{z}_i \right) \\
\leq\ & \sum_{l=1}^{L_n} \sum_{k=1}^{K_n} \exp\left(\frac{-\epsilon^2/4}{C\sqrt{q_n} n^{-1/2} + \epsilon C\sqrt{q_n} n^{-1/2}} \right) \\
\leq\ & \sum_{l=1}^{L_n} \sum_{k=1}^{K_n} \exp\left(-C\sqrt{n} q_n^{-1/2} \right) \\
\leq\ & C\left(C\sqrt{n d_n} \right)^{q_n} \left(C\sqrt{n} d_n \right)^{J_n} \exp\left(-C\sqrt{n} q_n^{-1/2} \right) \\
\leq\ & C\exp\left(C\left(d_n \log n - \sqrt{n} q_n^{-1/2} \right) \right) \to 0,
\end{aligned}
$$

by Condition 3 and 4. Note that the upper bound does not depend on $\{\mathbf{x}_i, \mathbf{z}_i\}$, so the above bound also holds unconditionally. This implies (B.5). The proof is finished. □

Proof of Lemma 6.

PROOF. It is sufficient to show that for any positive constants M and C,

$$P\left(\inf_{||\boldsymbol{\theta}_1-\tilde{\boldsymbol{\theta}}_1||\geq M,\,||\boldsymbol{\theta}_2||\leq C\sqrt{d_n}}\sum_{i=1}^n \tilde{Q}_i(\boldsymbol{\theta}_1,\tilde{\boldsymbol{\theta}}_1,\boldsymbol{\theta}_2)>0\right)\to 1. \tag{B.6}$$

By Lemma B.4,

$$\sup_{\substack{||\boldsymbol{\theta}_1-\tilde{\boldsymbol{\theta}}_1||\leq M\\ ||\boldsymbol{\theta}_2||\leq C\sqrt{d_n}}}\left|\sum_{i=1}^n \tilde{Q}_i(\boldsymbol{\theta}_1,\tilde{\boldsymbol{\theta}}_1,\boldsymbol{\theta}_2)-E_s\left[\tilde{Q}_i(\boldsymbol{\theta}_1,\tilde{\boldsymbol{\theta}}_1,\boldsymbol{\theta}_2)\right]+\tilde{\mathbf{x}}_i'\left(\boldsymbol{\theta}_1-\tilde{\boldsymbol{\theta}}_1\right)\psi_\tau(\epsilon_i)\right|=o_p(1).$$

Then by Lemma B.3,

$$\begin{aligned}\sup_{\substack{||\boldsymbol{\theta}_1-\tilde{\boldsymbol{\theta}}_1||\leq M\\ ||\boldsymbol{\theta}_2||\leq C\sqrt{d_n}}}\Bigg|&\sum_{i=1}^n\left[\tilde{Q}_i(\boldsymbol{\theta}_1,\tilde{\boldsymbol{\theta}}_1,\boldsymbol{\theta}_2)+\tilde{\mathbf{x}}_i'\left(\boldsymbol{\theta}_1-\tilde{\boldsymbol{\theta}}_1\right)\psi_\tau(\epsilon_i)\right]\\ &-\frac{1}{2}\left(\boldsymbol{\theta}_1'K_n\boldsymbol{\theta}_1-\tilde{\boldsymbol{\theta}}_1'K_n\tilde{\boldsymbol{\theta}}_1\right)(1+o_p(1))\Bigg|=o_p(1).\end{aligned} \tag{B.7}$$

Notice it follows from the derivation of Lemma 5 that

$$\begin{aligned}\left(\boldsymbol{\theta}_1-\tilde{\boldsymbol{\theta}}_1\right)'\sum_{i=1}^n\tilde{\mathbf{x}}_i\psi_\tau(\epsilon_i) &= \left(\boldsymbol{\theta}_1-\tilde{\boldsymbol{\theta}}_1\right)'n^{-1/2}X^{*\prime}\psi_\tau(\epsilon)\\ &= \left(\boldsymbol{\theta}_1-\tilde{\boldsymbol{\theta}}_1\right)'K_n\tilde{\boldsymbol{\theta}}_1(1+o_p(1)).\end{aligned} \tag{B.8}$$

Combining (B.7) and (B.8), we obtain

$$\begin{aligned}\sup_{\substack{||\boldsymbol{\theta}_1-\tilde{\boldsymbol{\theta}}_1||\leq M\\ ||\boldsymbol{\theta}_2||\leq C\sqrt{d_n}}}\Bigg|&\sum_{i=1}^n\left[\tilde{Q}_i(\boldsymbol{\theta}_1,\tilde{\boldsymbol{\theta}}_1,\boldsymbol{\theta}_2)\right]+\left(\boldsymbol{\theta}_1-\tilde{\boldsymbol{\theta}}_1\right)'K_n\tilde{\boldsymbol{\theta}}_1(1+o_p(1))\\ &-\frac{1}{2}\left(\boldsymbol{\theta}_1'K_n\boldsymbol{\theta}_1-\tilde{\boldsymbol{\theta}}_1'K_n\tilde{\boldsymbol{\theta}}_1\right)\Bigg|=o_p(1),\end{aligned}$$

which implies,

$$\sup_{\substack{||\boldsymbol{\theta}_1-\tilde{\boldsymbol{\theta}}_1||\leq M\\ ||\boldsymbol{\theta}_2||\leq C\sqrt{d_n}}}\left|\sum_{i=1}^n\left[\tilde{Q}_i(\boldsymbol{\theta}_1,\tilde{\boldsymbol{\theta}}_1,\boldsymbol{\theta}_2)\right]-\frac{1}{2}\left(\boldsymbol{\theta}_1-\tilde{\boldsymbol{\theta}}_1\right)'K_n\left(\boldsymbol{\theta}_1-\tilde{\boldsymbol{\theta}}_1\right)(1+o_p(1))\right|=o_p(1).$$

By Condition 2, for any $||\boldsymbol{\theta}_1 - \tilde{\boldsymbol{\theta}}_1|| > M$,

$$\frac{1}{2}\left(\boldsymbol{\theta}_1 - \tilde{\boldsymbol{\theta}}_1\right)' K_n \left(\boldsymbol{\theta}_1 - \tilde{\boldsymbol{\theta}}_1\right) > CM,$$

for some positive constant C. Thus (B.6) holds. This finishes the proof. □

LEMMA B.5. *Assume Conditions 1-6 hold, $n^{-1/2}q_n = o(\lambda)$, $n^{-1/2}k_n = o(\lambda)$ and $\log(p_n) = o(n\lambda^2)$. Then for some positive constant C,*

$$P\left(\max_{q_n+1\le j\le p_n} \sup_{\substack{||\boldsymbol{\beta}_1-\boldsymbol{\beta}_{01}||\le C\sqrt{\frac{q_n}{n}} \\ ||\boldsymbol{\gamma}-\boldsymbol{\gamma}_0||\le C\sqrt{\frac{d_n k_n}{n}}}} \left|\sum_{i=1}^n x_{ij}\Big[I(Y_i - \mathbf{x}_{A_i}{}'\boldsymbol{\beta}_1 - \mathbf{W}(\mathbf{z}_i)'\boldsymbol{\gamma} \le 0) \right.\right.$$
$$-I(Y_i - \mathbf{x}_{A_i}{}'\boldsymbol{\beta}_{01} - g_0(\mathbf{z}_i) \le 0) - P(Y_i - \mathbf{x}_{A_i}{}'\boldsymbol{\beta}_1 - \mathbf{W}(\mathbf{z}_i)'\boldsymbol{\gamma} \le 0)$$
$$\left.\left. +P(Y_i - \mathbf{x}_{A_i}{}'\boldsymbol{\beta}_{01} - g_0(\mathbf{z}_i) \le 0)\Big]\right| > n\lambda\right) \to 0.$$

PROOF. Extending Welsh (1989), we consider $\mathcal{B} = \left\{\boldsymbol{\beta}_1 : ||\boldsymbol{\beta}_1 - \boldsymbol{\beta}_{01}|| \le C\sqrt{\frac{q_n}{n}}\right\}$ and $\mathcal{G} = \left\{\boldsymbol{\gamma} : ||\boldsymbol{\gamma} - \boldsymbol{\gamma}_0|| \le C\sqrt{\frac{k_n d_n}{n}}\right\}$. The sets of $\mathcal{B}$ and $\mathcal{G}$ can be covered with a net of balls with radius $C\sqrt{q_n/n^5}$ and $C\sqrt{k_n d_n/n^5}$, respectively, with cardinality $N_1 \equiv |\mathcal{B}| \le Cn^{2q_n}$ and $N_2 \equiv |\mathcal{G}| \le Cn^{2k_n d_n}$. Denote the N_1 balls by $\boldsymbol{\beta}(\mathbf{t}_1), ..., \boldsymbol{\beta}(\mathbf{t}_{N_1})$, where the ball $\boldsymbol{\beta}(\mathbf{t}_k)$ is centered at $\mathbf{t}_k$, $k = 1, ..., N_1$; and denote the N_2 balls by $\boldsymbol{\gamma}(\mathbf{u}_1), ..., \boldsymbol{\gamma}(\mathbf{u}_{N_2})$, where the ball $\boldsymbol{\gamma}(\mathbf{u}_k)$ is centered at $\mathbf{u}_k$, $k = 1, ..., N_2$. Let $\epsilon_i(\boldsymbol{\beta}_1, \boldsymbol{\gamma}) = Y_i - \mathbf{x}_{A_i}{}'\boldsymbol{\beta}_1 - \mathbf{W}(\mathbf{z}_i)'\boldsymbol{\gamma}$ and $\epsilon_i = Y_i - \mathbf{x}_{A_i}{}'\boldsymbol{\beta}_{01} - g_0(\mathbf{z}_i)$.

$$P\left(\sup_{\substack{||\boldsymbol{\beta}_1-\boldsymbol{\beta}_{01}||\le C\sqrt{\frac{q_n}{n}} \\ ||\boldsymbol{\gamma}-\boldsymbol{\gamma}_0||\le C\sqrt{\frac{k_n d_n}{n}}}} \left|\sum_{i=1}^n x_{ij}\Big[I(\epsilon_i(\boldsymbol{\beta}_1, \boldsymbol{\gamma}) \le 0) - I(\epsilon_i \le 0) - P(\epsilon_i(\boldsymbol{\beta}_1, \boldsymbol{\gamma}) \le 0)\right.\right.$$
$$\left.\left. +P(\epsilon_i \le 0)\Big]\right| > n\lambda\right)$$
$$\le \sum_{l=1}^{N_2}\sum_{k=1}^{N_1} P\left(\left|\sum_{i=1}^n x_{ij}\Big[I(\epsilon_i(\mathbf{t}_k, \mathbf{u}_l) \le 0) - I(\epsilon_i \le 0) - P\left(\epsilon_i(\mathbf{t}_k, \mathbf{u}_l) \le 0\right) + P\left(\epsilon_i \le 0\right)\Big]\right| > n\lambda/2\right)$$

$$
\begin{aligned}
& + \sum_{l=1}^{N_2}\sum_{k=1}^{N_1} P\Bigg(\sup_{\substack{||\tilde{\boldsymbol{\beta}}_1-\mathbf{t}_k||\le C\sqrt{q_n/n^5} \\ ||\tilde{\boldsymbol{\gamma}}-\mathbf{u}_l||\le C\sqrt{k_n d_n/n^5}}} \Bigg| \sum_{i=1}^n x_{ij}\Big[I\left(\epsilon_i(\tilde{\boldsymbol{\beta}}_1,\tilde{\boldsymbol{\gamma}})\le 0\right) - I\left(\epsilon_i(\mathbf{t}_k,\mathbf{u}_l)\le 0\right) \\
& -P\left(\epsilon_i(\tilde{\boldsymbol{\beta}}_1,\tilde{\boldsymbol{\gamma}})\le 0\right) + P\left(\epsilon_i(\mathbf{t}_k,\mathbf{u}_l)\le 0\right)\Big]\Bigg| > n\lambda/2\Bigg) \\
\equiv\ & I_{nj1} + I_{nj2}.
\end{aligned}
$$

To evaluate I_{nj1}, let $\nu_{ij} = x_{ij}\left[I(\epsilon_i(\mathbf{t}_k,\mathbf{u}_l)\le 0) - I(\epsilon_i \le 0) - P\left(\epsilon_i(\mathbf{t}_k,\mathbf{u}_l)\le 0\right) + P\left(\epsilon_i\le 0\right)\right]$, which are bounded, independent mean-zero random variables. Note that $\mathrm{Var}\left(\nu_{ij}\right) = \mathrm{Var}\big(x_{ij}\left[I(\epsilon_i(\mathbf{t}_k,\mathbf{u}_l)\le 0) - I(\epsilon_i\le 0)\right]\big)$. Since $\epsilon_i(\mathbf{t}_k,\mathbf{u}_l) = \epsilon_i - (\mathbf{x}_{A_i}{}'(\mathbf{t}_k-\boldsymbol{\beta}_{01}) + \mathbf{W}(\mathbf{z}_i)'(\mathbf{u}_l-\boldsymbol{\gamma}_0) + u_{ni})$, $I(\epsilon_i(\mathbf{t}_k,\mathbf{u}_l)\le 0) - I(\epsilon_i\le 0)$ is nonzero only if $\epsilon_i \in (0, \mathbf{x}_{A_i}{}'(\mathbf{t}_k-\boldsymbol{\beta}_{01}) + \mathbf{W}(\mathbf{z}_i)'(\mathbf{u}_l-\boldsymbol{\gamma}_0) + u_{ni})$ or $\epsilon_i \in (\mathbf{x}_{A_i}{}'(\mathbf{t}_k-\boldsymbol{\beta}_{01}) + \mathbf{W}(\mathbf{z}_i)'(\mathbf{u}_l-\boldsymbol{\gamma}_0) + u_{ni}, 0)$. Hence,

$$
\begin{aligned}
\mathrm{Var}\left(\nu_{ij}\right) &\le \mathrm{E}x_{ij}^2\left[I(\epsilon_i(\mathbf{t}_k,\mathbf{u}_l)\le 0) - I(\epsilon_i\le 0)\right]^2\big) \\
&\le CP(|\epsilon_i| < |\mathbf{x}_{A_i}{}'(\mathbf{t}_k-\boldsymbol{\beta}_{01}) + \mathbf{W}(\mathbf{z}_i)'(\mathbf{u}_l-\boldsymbol{\gamma}_0) + u_{ni}|) \\
&\quad \big(1 - P(|\epsilon_i| < |\mathbf{x}_{A_i}{}'(\mathbf{t}_k-\boldsymbol{\beta}_{01}) + \mathbf{W}(\mathbf{z}_i)'(\mathbf{u}_l-\boldsymbol{\gamma}_0) + u_{ni}|)\big) \\
&\le C\big[F_i(|\mathbf{x}_{A_i}{}'(\mathbf{t}_k-\boldsymbol{\beta}_{01}) + \mathbf{W}(\mathbf{z}_i)'(\mathbf{u}_l-\boldsymbol{\gamma}_0) + u_{ni}|) - F_i(0)\big] \\
&\le C|\mathbf{x}_{A_i}{}'(\mathbf{t}_k-\boldsymbol{\beta}_{01}) + \mathbf{W}(\mathbf{z}_i)'(\mathbf{u}_l-\boldsymbol{\gamma}_0) + u_{ni}|.
\end{aligned}
$$

Therefore, by Jensen's inequality,

$$
\begin{aligned}
& \sum_{i=1}^n \mathrm{Var}(\nu_{ij}) \\
\le\ & C\sum_{i=1}^n |\mathbf{x}_{A_i}{}'(\mathbf{t}_k-\boldsymbol{\beta}_{01}) + \mathbf{W}(\mathbf{z}_i)'(\mathbf{u}_l-\boldsymbol{\gamma}_0) + u_{ni}| \\
\le\ & Cn\big[n^{-1}\sum_i \left(\mathbf{x}_{A_i}{}'(\mathbf{t}_k-\boldsymbol{\beta}_{01}) + \mathbf{W}(\mathbf{z}_i)'(\mathbf{u}_l-\boldsymbol{\gamma}_0) + u_{ni}\right)^2\big]^{1/2} \\
\le\ & C\big[n\lambda_{\max}^{1/2}(n^{-1}X_A X_A')||\mathbf{t}_k-\boldsymbol{\beta}_{01}|| + n\sqrt{n^{-1}(\mathbf{u}_l-\boldsymbol{\gamma}_0)'W^2(\mathbf{u}_l-\boldsymbol{\gamma}_0)} \\
& +n\sup_i |u_{ni}|\big] \\
\le\ & C\left(\sqrt{nq_n} + \sqrt{nd_n} + nk_n^{-r}\right) \le C\sqrt{nd_n}.
\end{aligned}
$$

Applying Bernstein's inequality,

$$\begin{aligned} I_{nj1} &\leq N_1 N_2 \exp\left(-\frac{n^2\lambda^2/8}{C\sqrt{nd_n}+Cn\lambda}\right) \\ &\leq N_1 N_2 \exp\left(-Cn\lambda\right) \\ &= C\exp\left(Cq_n\log(n)+Cd_nk_n\log(n)-Cn\lambda\right). \end{aligned}$$

To evaluate I_{nj2}, note that $I(\epsilon_i(\tilde{\boldsymbol{\beta}},\tilde{\boldsymbol{\gamma}})\leq 0) = I\left(\epsilon_i(\mathbf{t}_k,\mathbf{u}_l)\leq \mathbf{x}_{A_i}{}'(\tilde{\boldsymbol{\beta}}-\mathbf{t}_k)+\mathbf{W}(\mathbf{z}_i)'(\tilde{\boldsymbol{\gamma}}-\mathbf{u}_l)\right)$. Since $I(x\leq s)$ is an increasing function of s, we have

$$\begin{aligned} &\sup_{\substack{\left|\left|\tilde{\boldsymbol{\beta}}_1-\mathbf{t}_k\right|\right|\leq C\sqrt{q_n/n^5} \\ \left|\left|\tilde{\boldsymbol{\gamma}}-\mathbf{u}_l\right|\right|\leq C\sqrt{k_nd_n/n^5}}} \left|\sum_{i=1}^n x_{ij}\left[I\left(\epsilon_i\left(\tilde{\boldsymbol{\beta}}_1,\tilde{\boldsymbol{\gamma}}\right)\leq 0\right)-I\left(\epsilon_i\left(\mathbf{t}_k,\mathbf{u}_l\right)\leq 0\right)-P\left(\epsilon_i\left(\tilde{\boldsymbol{\beta}}_1,\tilde{\boldsymbol{\gamma}}\right)\leq 0\right)\right.\right. \\ &\left.\left.+P\left(\epsilon_i\left(\mathbf{t}_k,\mathbf{u}_l\right)\leq 0\right)\right]\right| \\ \leq\; &\sum_{i=1}^n |x_{ij}|\left[I\left(\epsilon_i\left(\mathbf{t}_k,\mathbf{u}_l\right)\leq C\sqrt{q_n/n^5}||\mathbf{x}_{A_i}||+C\sqrt{k_nd_n/n^5}||\mathbf{W}(\mathbf{z}_i)||\right)-I\left(\epsilon_i\left(\mathbf{t}_k,\mathbf{u}_l\right)\leq 0\right)\right. \\ &\left.-P\left(\epsilon_i\left(\mathbf{t}_k,\mathbf{u}_l\right)\leq -C\sqrt{q_n/n^5}||\mathbf{x}_{A_i}||-C\sqrt{k_nd_n/n^5}||\mathbf{W}(\mathbf{z}_i)||\right)+P\left(\epsilon_i\left(\mathbf{t}_k,\mathbf{u}_l\right)\leq 0\right)\right] \\ =\; &\sum_{i=1}^n |x_{ij}|\left[I\left(\epsilon_i\left(\mathbf{t}_k,\mathbf{u}_l\right)\leq C\sqrt{q_n/n^5}||\mathbf{x}_{A_i}||+C||\mathbf{W}(\mathbf{z}_i)||\sqrt{k_nd_n/n^5}\right)-I\left(\epsilon_i\left(\mathbf{t}_k,\mathbf{u}_l\right)\leq 0\right)\right. \\ &\left.-P\left(\epsilon_i\left(\mathbf{t}_k,\mathbf{u}_l\right)\leq C\sqrt{q_n/n^5}||\mathbf{x}_{A_i}||+C||\mathbf{W}(\mathbf{z}_i)||\sqrt{k_nd_n/n^5}\right)+P\left(\epsilon_i\left(\mathbf{t}_k,\mathbf{u}_l\right)\leq 0\right)\right] \\ &+\sum_{i=1}^n |x_{ij}|\left[P\left(\epsilon_i\left(\mathbf{t}_k,\mathbf{u}_l\right)\leq C\sqrt{q_n/n^5}||\mathbf{x}_{A_i}||+C||\mathbf{W}(\mathbf{z}_i)||\sqrt{k_nd_n/n^5}\right)\right. \\ &\left.-P\left(\epsilon_i\left(\mathbf{t}_k,\mathbf{u}_l\right)\leq -C\sqrt{q_n/n^5}||\mathbf{x}_{A_i}||-C\sqrt{k_nd_n/n^5}||\mathbf{W}(\mathbf{z}_i)||\right)\right]. \end{aligned}$$

Note that

$$\begin{aligned}
& \sum_{i=1}^{n}|x_{ij}|\Big[P\left(\epsilon_i\left(\mathbf{t}_k,\mathbf{u}_l\right)\leq C\sqrt{q_n/n^5}||\mathbf{x}_{A_i}||+C\sqrt{k_nd_n/n^5}||\mathbf{W}(\mathbf{z}_i)||\right)\\
& -P\left(\epsilon_i\left(\mathbf{t}_k,\mathbf{u}_l\right)\leq -C\sqrt{q_n/n^5}||\mathbf{x}_{A_i}||-C\sqrt{k_nd_n/n^5}||\mathbf{W}(\mathbf{z}_i)||\right)\Big]\\
= & \sum_{i=1}^{n}|x_{ij}|\Big[F_i\left(C\sqrt{q_n/n^5}||\mathbf{x}_{A_i}||+C||\mathbf{W}(\mathbf{z}_i)||\sqrt{k_nd_n/n^5}+\mathbf{x}_{A_i}{}'(\mathbf{t}_k-\boldsymbol{\beta}_{01})+\mathbf{W}(\mathbf{z}_i)'(\boldsymbol{\gamma}-\boldsymbol{\gamma}_0)-u_{ni}\right)\\
& -F_i\left(-C\sqrt{q_n/n^5}||\mathbf{x}_{A_i}||-C||\mathbf{W}(\mathbf{z}_i)||\sqrt{k_nd_n/n^5}+\mathbf{x}_{A_i}{}'(\mathbf{t}_k-\boldsymbol{\beta}_{01})+\mathbf{W}(\mathbf{z}_i)'(\boldsymbol{\gamma}-\boldsymbol{\gamma}_0)-u_{ni}\right)\Big]\\
\leq & \; C\sum_{i=1}^{n}\left(\sqrt{q_n/n^5}||\mathbf{x}_{A_i}||+||\mathbf{W}(\mathbf{z}_i)||\sqrt{k_nd_n/n^5}\right)\leq Cd_nn^{-3/2}=o(n\lambda).
\end{aligned}$$

Hence, for all n sufficiently large, $I_{nj2}\leq\sum_{l=1}^{N_2}\sum_{k=1}^{N_1}P\left(\sum_{i=1}^{n}\alpha_{lki}\geq\frac{n\lambda}{4}\right)$,

where

$$\begin{aligned}
\alpha_{lki} = & \; |x_{ij}|\Big[I\left(\epsilon_i\left(\mathbf{t}_k,\mathbf{u}_l\right)\leq C\sqrt{q_n/n^5}||\mathbf{x}_{A_i}||+C||\mathbf{W}(\mathbf{z}_i)||\sqrt{k_nd_n/n^5}\right)-I\left(\epsilon_i\left(\mathbf{t}_k,\mathbf{u}_l\right)\leq 0\right)\\
& - P\left(\epsilon_i\left(\mathbf{t}_k,\mathbf{u}_l\right)\leq C\sqrt{q_n/n^5}||\mathbf{x}_{A_i}||+C||\mathbf{W}(\mathbf{z}_i)||\sqrt{k_nd_n/n^5}\right)+P\left(\epsilon_i\left(\mathbf{t}_k,\mathbf{u}_l\right)\leq 0\right)\Big].
\end{aligned}$$

Note that by Condition 2, α_{lki} are independent bounded random variables with mean zero. Similarly as in the evaluation of I_{nj2}, we can show that

$$\mathrm{Var}(\alpha_{lki})\leq C\left(\sqrt{q_n/n^5}||\mathbf{x}_{A_i}||+||\mathbf{W}(\mathbf{z}_i)||\sqrt{k_nd_n/n^5}\right)<Cd_nn^{-5/2}.$$

Applying Bernstein's inequality,

$$\begin{aligned}
\sum_{l=1}^{N_2}\sum_{k=1}^{N_1}P\left(\sum_{i=1}^{n}\alpha_{lki}\geq\frac{n\lambda}{4}\right) & \leq N_1N_2\mathrm{exp}\left(-\frac{n^2\lambda^2/32}{Cd_nn^{-3/2}+Cn\lambda}\right)\\
& \leq N_1N_2\mathrm{exp}(-Cn\lambda)\\
& \leq C\mathrm{exp}((Cq_n\log(n)+Cd_nk_n\log(n)-Cn\lambda).
\end{aligned}$$

Therefore, for all n sufficiently large, the probability of interest in the lemma is bounded by

$$\begin{aligned}
\sum_{j=q_n+1}^{p_n}(I_{nj1}+I_{nj2}) & \leq C\mathrm{exp}(\log(p_n)+(Cq_n\log(n)+Cd_nk_n\log(n)-Cn\lambda)\\
& = o(1).
\end{aligned}$$

This completes the proof. □

Proof of Lemma 1 (3.5).

PROOF. Let $\mathcal{D} = \{i : Y_i - \mathbf{x}_{A_i}{}'\hat{\boldsymbol\beta}_1 - \boldsymbol\Pi(\mathbf{z}_i)'\hat{\boldsymbol\xi} = 0\}$, then for $j = q_n+1, ..., p_n$

$$s_j\left(\hat{\boldsymbol\beta}, \hat{\boldsymbol\xi}\right) = \frac{1}{n}\sum_{i=1}^{n} x_{ij}\left[I\left(Y_i - \mathbf{x}_{A_i}{}'\hat{\boldsymbol\beta}_1 - \boldsymbol\Pi(\mathbf{z}_i)'\hat{\boldsymbol\xi} \le 0\right) - \tau\right] - \frac{1}{n}\sum_{i\in\mathcal{D}} x_{ij}(a_i^* + (1-\tau)),$$

where a_i^* satisfy the condition of Lemma 1. With probability one (Section 2.2, Koenker, 2005), $|\mathcal{D}| = d_n$. Therefore, $n^{-1}\sum_{i\in\mathcal{D}} x_{ij}(a_i^* + (1-\tau)) = O_p\left(d_n n^{-1}\right) = o_p(\lambda)$. We will show that

(B.9)
$$P\left(\max_{q_n+1\le j\le p_n}\left|n^{-1}\sum_{i=1}^{n} x_{ij}\left[I(Y_i - \mathbf{x}_{A_i}{}'\hat{\boldsymbol\beta}_1 - \boldsymbol\Pi(\mathbf{z}_i)'\hat{\boldsymbol\xi} \le 0) - \tau\right]\right| > c\lambda\right) \to 0.$$

Note that at the oracle estimator, is it equivalent to showing

$$P\left(\max_{q_n+1\le j\le p_n}\left|n^{-1}\sum_{i=1}^{n} x_{ij}\left[I(Y_i - \mathbf{x}_{A_i}{}'\hat{\boldsymbol\beta}_1 - \mathbf{W}(\mathbf{z}_i)'\hat{\boldsymbol\gamma} \le 0) - \tau\right]\right| > c\lambda\right) \to 0.$$

It follows from the proof of Lemmas 2 and 4 that $||\hat{\boldsymbol\gamma} - \boldsymbol\gamma_0|| = O_p\left(\sqrt{\frac{k_n d_n}{n}}\right)$.

Note that

$$\begin{aligned}
& P\left(\max_{q_n+1\le j\le p_n}\left|n^{-1}\sum_{i=1}^{n} x_{ij}\left[I(Y_i - \mathbf{x}_{A_i}{}'\hat{\boldsymbol\beta}_1 - \mathbf{W}(\mathbf{z}_i)'\hat{\boldsymbol\gamma} \le 0) - \tau\right]\right| > c\lambda\right) \\
\le\ & P\Big(\max_{q_n+1\le j\le p_n}\left|n^{-1}\sum_{i=1}^{n} x_{ij}\left[I(Y_i - \mathbf{x}_{A_i}{}'\hat{\boldsymbol\beta}_1 - \mathbf{W}(\mathbf{z}_i)'\hat{\boldsymbol\gamma} \le 0) - I(Y_i - \mathbf{x}_{A_i}{}'\boldsymbol\beta_{01} - g_0(\mathbf{z}_i) \le 0)\right]\right| \\
& > c\lambda/2\Big) + P\left(\max_{q_n+1\le j\le p_n}\left|n^{-1}\sum_{i=1}^{n} x_{ij}\left[I(Y_i - \mathbf{x}_{A_i}{}'\boldsymbol\beta_{01} - g_0(\mathbf{z}_i) \le 0) - \tau\right]\right| > c\lambda/2\right) \\
\le\ & P\Bigg(\max_{q_n+1\le j\le p_n}\sup_{\substack{||\boldsymbol\beta_1-\boldsymbol\beta_{01}||\le Cq_n^{1/2}n^{-1/2} \\ ||\boldsymbol\gamma-\boldsymbol\gamma_0||\le C\sqrt{\frac{k_n d_n}{n}}}}\Bigg|n^{-1}\sum_{i=1}^{n} x_{ij}\Big[I(Y_i - \mathbf{x}_{A_i}{}'\boldsymbol\beta_1 - \mathbf{W}(\mathbf{z}_i)'\boldsymbol\gamma \le 0) \\
& -I(Y_i - \mathbf{x}_{A_i}{}'\boldsymbol\beta_{01} - g_0(\mathbf{z}_i) \le 0)\Big]\Bigg| > c\lambda/2\Bigg) + o_p(1)
\end{aligned}$$

$$
\begin{aligned}
\leq\ & P\Bigg(\max_{q_n+1\leq j\leq p_n}\sup_{\substack{||\boldsymbol{\beta}_1-\boldsymbol{\beta}_{01}||\leq Cq_n^{1/2}n^{-1/2}\\ ||\boldsymbol{\gamma}-\boldsymbol{\gamma}_0||\leq C\sqrt{\frac{k_nd_n}{n}}}}\Big|n^{-1}\sum_{i=1}^n x_{ij}\Big[I(Y_i-\mathbf{x}_{A_i}{}'\boldsymbol{\beta}_1-\mathbf{W}(\mathbf{z}_i)'\boldsymbol{\gamma}\leq 0)\\
&-I(Y_i-\mathbf{x}_{A_i}{}'\boldsymbol{\beta}_{01}-g_0(\mathbf{z}_i)\leq 0)-P(Y_i-\mathbf{x}_{A_i}{}'\boldsymbol{\beta}_1-\mathbf{W}(\mathbf{z}_i)'\boldsymbol{\gamma}\leq 0)\\
&+P(Y_i-\mathbf{x}_{A_i}{}'\boldsymbol{\beta}_{01}-g_0(\mathbf{z}_i)\leq 0)\Big]\Big|>c\lambda/4\Bigg)\\
&+P\Bigg(\max_{q_n+1\leq j\leq p_n}\sup_{\substack{||\boldsymbol{\beta}_1-\boldsymbol{\beta}_{01}||\leq Cq_n^{1/2}n^{-1/2}\\ ||\boldsymbol{\gamma}-\boldsymbol{\gamma}_0||\leq C\sqrt{\frac{k_nd_n}{n}}}}\Big|n^{-1}\sum_{i=1}^n x_{ij}\Big[P(Y_i-\mathbf{x}_{A_i}{}'\boldsymbol{\beta}_1-\mathbf{W}(\mathbf{z}_i)'\boldsymbol{\gamma}\leq 0)\\
&-P(Y_i-\mathbf{x}_{A_i}{}'\boldsymbol{\beta}_{01}-g_0(\mathbf{z}_i)\leq 0)\Big]\Big|>c\lambda/4\Bigg)+o_p(1),
\end{aligned}
$$

where the second inequality applies Lemma B.2. Note that

$$
\begin{aligned}
&\max_{q_n+1\leq j\leq p_n}\sup_{\substack{||\boldsymbol{\beta}_1-\boldsymbol{\beta}_{01}||\leq Cq^{1/2}n^{-1/2}\\ ||\boldsymbol{\gamma}-\boldsymbol{\gamma}_0||\leq C\sqrt{\frac{k_nd_n}{n}}}}\Big|n^{-1}\sum_{i=1}^n x_{ij}\Big[P(Y_i-\mathbf{x}_{A_i}{}'\boldsymbol{\beta}_1-\mathbf{W}(\mathbf{z}_i)'\boldsymbol{\gamma}\leq 0)\\
&-P(Y_i-\mathbf{x}_{A_i}{}'\boldsymbol{\beta}_{01}-g_0(\mathbf{z}_i)\leq 0)\Big]\Big|\\
=\ &\max_{q_n+1\leq j\leq p_n}\sup_{\substack{||\boldsymbol{\beta}_1-\boldsymbol{\beta}_{01}||\leq Cq^{1/2}n^{-1/2}\\ ||\boldsymbol{\gamma}-\boldsymbol{\gamma}_0||\leq C_2\sqrt{\frac{k_nd_n}{n}}}}\Big|n^{-1}\sum_{i=1}^n x_{ij}\big[F_i(\mathbf{x}_{A_i}{}'(\boldsymbol{\beta}_1-\boldsymbol{\beta}_{01})+\mathbf{W}(\mathbf{z}_i)'(\boldsymbol{\gamma}-\boldsymbol{\gamma}_0)\\
&-u_{ni})-F_i(0)\big]\Big|\\
\leq\ & C\sup_{\substack{||\boldsymbol{\beta}_1-\boldsymbol{\beta}_{01}||\leq Cq^{1/2}n^{-1/2}\\ ||\boldsymbol{\gamma}-\boldsymbol{\gamma}_0||\leq C\sqrt{\frac{d_nk_n}{n}}}} n^{-1}\sum_{i=1}^n\big(|\mathbf{x}_{A_i}{}'(\boldsymbol{\beta}_1-\boldsymbol{\beta}_{01})|+|\mathbf{W}(\mathbf{z}_i)'(\boldsymbol{\gamma}-\boldsymbol{\gamma}_0)|+|u_{ni}|\big)\\
\leq\ & C\sup_{\substack{||\boldsymbol{\beta}_1-\boldsymbol{\beta}_{01}||\leq Cq^{1/2}n^{-1/2}\\ ||\boldsymbol{\gamma}-\boldsymbol{\gamma}_0||\leq C\sqrt{\frac{d_nk_n}{n}}}}\big[\lambda_{\max}^{1/2}(n^{-1}X_AX_A')||\boldsymbol{\beta}_1-\boldsymbol{\beta}_{01}||+\sqrt{n^{-1}(\boldsymbol{\gamma}-\boldsymbol{\gamma}_0)'W^2(\boldsymbol{\gamma}-\boldsymbol{\gamma}_0)}\\
&+\sup_i|u_{ni}|\big]\\
\leq\ & C\left(q_n^{1/2}n^{-1/2}+\sqrt{d_n/n}+k_n^{-r}\right)=o(\lambda),
\end{aligned}
$$

where the second last inequality applies Jensen's inequality (see the proof of Lemma B.5 for more details).

Hence, Lemma 1 (3.5) follows if we can show

$$P\Bigg(\max_{q_n+1\le j\le p_n}\sup_{\substack{||\boldsymbol{\beta}_1-\boldsymbol{\beta}_{01}||\le Cq_n^{1/2}n^{-1/2}\\ ||\boldsymbol{\gamma}-\boldsymbol{\gamma}_0||\le C\sqrt{\frac{k_nd_n}{n}}}}\Bigg|n^{-1}\sum_{i=1}^{n}x_{ij}\Big[I(Y_i-\mathbf{x}_{A_i}{}'\boldsymbol{\beta}_1-\mathbf{W}(\mathbf{z}_i)'\boldsymbol{\gamma}\le 0)$$
$$-I(Y_i-\mathbf{x}_{A_i}{}'\boldsymbol{\beta}_{01}-g_0(\mathbf{z}_i)\le 0)-P(Y_i-\mathbf{x}_{A_i}{}'\boldsymbol{\beta}_1-\mathbf{W}(\mathbf{z}_i)'\boldsymbol{\gamma}\le 0)$$
$$+P(Y_i-\mathbf{x}_{A_i}{}'\boldsymbol{\beta}_{01}-g_0(\mathbf{z}_i)\le 0)\Big]\Bigg|>c\lambda/4\Bigg)\to 0.$$

The above statement holds by Lemma B.5. This proves the result. □

B.2. Technical details for results in Section 6 of the main paper. The following conditions are used for the derivation of Theorem 6.1 of the main paper. These conditions are similar as those in Section 3.

CONDITION B.1. *(Conditions on the random errors) The random error $\epsilon_i^{(m)}=Y_i-Q_{Y_i|\mathbf{x}_i,\mathbf{z}_i}(\tau_m)$ has the conditional distribution function $F_i^{(m)}$ and continuous conditional density function $f_i^{(m)}$, given $\mathbf{x}_i$, $\mathbf{z}_i$. The $f_i^{(m)}$ are uniformly bounded away from 0 and infinity in a neighborhood of zero, its first derivative $f_i'^{(m)}$ has a uniform upper bound in a neighborhood of zero, for $1\le i\le n$, $1\le m\le M$.*

CONDITION B.2. *(Conditions on the covariates) There exist positive constants M_1 and M_2 such that $|x_{ij}|\le M_1$, $\forall\ 1\le i\le n, 1\le j\le p_n$. Let $\delta_{ij}^{(m)}$ and $\Delta_n^{(m)}$ follow from the definitions in Section 2.2 for quantile τ_m. Assume $E\left[{\delta_{ij}^{(m)}}^4\right]\le M_2$, $\forall\ 1\le i\le n$, $1\le j\le\bar{q}_n$, $1\le m\le M$. There exist finite positive constants C_1 and C_2 such that with probability one*

$$C_1\le\lambda_{\max}\left(n^{-1}X_AX_A'\right)\le C_2,\quad C_1\le\lambda_{\max}\left(n^{-1}\Delta_n^{(m)}{\Delta_n^{(m)}}'\right)\le C_2,\ 1\le m\le M.$$

CONDITION B.3. *(Condition on the nonlinear functions for multiple quantiles) For $r=m+v>1.5$, $g_0^{(m)}\in\mathcal{G}$, $\forall\ 1\le m\le M$.*

CONDITION B.4. *(Condition on the B-spline basis functions) The dimension of the spline basis k_n satisfies $k_n\approx n^{1/(2r+1)}$.*

CONDITION B.5. *(Condition on the joint oracle model) $\bar{q}_n=O\left(n^{C_3}\right)$ for some $C_3<\frac{1}{3}$.*

Condition B.6. *(Condition on the signal across quantiles) There exist positive constants C_4 and C_5 such that $2C_3 < C_4 < 1$ and*

$$n^{(1-C_4)/2} \min_{1 \le j \le \bar{q}_n, 1 \le m \le M} \left|\beta_{0j}^{(m)}\right| \ge C_5.$$

For the proof of Theorem 6.1, we note that the objective function $\bar{Q}^P(\boldsymbol{\beta}, \boldsymbol{\xi})$ can be decomposed as the difference of two convex functions. Extending the notation we used for the single quantile case, we write $\bar{Q}^P(\boldsymbol{\beta}, \boldsymbol{\xi}) = \bar{k}(\boldsymbol{\beta}, \boldsymbol{\xi}) - \bar{l}(\boldsymbol{\beta}, \boldsymbol{\xi})$, where $\bar{k}(\boldsymbol{\beta}, \boldsymbol{\xi}) = n^{-1} \sum_{m=1}^{M} \sum_{i=1}^{n} \rho_{\tau_m}(Y_i - \mathbf{x}_i' \boldsymbol{\beta}^{(m)} - \boldsymbol{\Pi}(\mathbf{z}_i)' \boldsymbol{\xi}^{(m)}) + \lambda \sum_{j=1}^{p_n} \sum_{m=1}^{M} |\beta_j^{(m)}|$, $\bar{l}(\boldsymbol{\beta}, \boldsymbol{\xi}) = \sum_{j=1}^{p_n} L(||\bar{\boldsymbol{\beta}}^j||_1)$, and $L(\cdot)$ has the function form given in Section 3 of the main paper for the SCAD or MCP penalty functions. However, here $\bar{l}(\boldsymbol{\beta}, \boldsymbol{\xi})$ is convex but not differentiable everywhere due to the group level penalty.

We will first analyze the subgradient of the unpenalized objective function, which plays a pivotal part in the proof of the oracle property. Generalize the sub-gradient from Section 3 of the main paper to the multiple quantile case such that for $1 \le m \le M$,

$$\begin{aligned}
s_j^{(m)}(\boldsymbol{\beta}, \boldsymbol{\xi}) &= -n^{-1} \tau_m \sum_{i=1}^{n} x_{ij} I(Y_i - \mathbf{x}_i' \boldsymbol{\beta}^{(m)} - \boldsymbol{\Pi}(\mathbf{z}_i)' \boldsymbol{\xi}^{(m)} > 0) \\
&\quad + n^{-1}(1 - \tau_m) \sum_{i=1}^{n} x_{ij} I(Y_i - \mathbf{x}_i' \boldsymbol{\beta}^{(m)} - \boldsymbol{\Pi}(\mathbf{z}_i)' \boldsymbol{\xi}^{(m)} < 0) \\
&\quad - n^{-1} \sum_{i=1}^{n} x_{ij} a_i^{(m)} \quad \text{for } 1 \le j \le p_n, \\
s_j^{(m)}(\boldsymbol{\beta}, \boldsymbol{\xi}) &= -n^{-1} \tau_m \sum_{i=1}^{n} \Pi_{j-p_n}(\mathbf{z}_i) I(Y_i - \mathbf{x}_i' \boldsymbol{\beta}^{(m)} - \boldsymbol{\Pi}(\mathbf{z}_i)' \boldsymbol{\xi}^{(m)} > 0) \\
&\quad + n^{-1}(1 - \tau_m) \sum_{i=1}^{n} \Pi_{j-p_n}(\mathbf{z}_i) I(Y_i - \mathbf{x}_i' \boldsymbol{\beta}^{(m)} - \boldsymbol{\Pi}(\mathbf{z}_i)' \boldsymbol{\xi}^{(m)} < 0) \\
&\quad - n^{-1} \sum_{i=1}^{n} \Pi_{j-p_n}(\mathbf{z}_i) a_i^{(m)} \quad \text{for } p_n + 1 \le j \le p_n + L_n,
\end{aligned}$$

where $a_i^{(m)}$ follows from definition of a_i in Section 3.3.

As in the single quantile case, understanding the asymptotic behavior of the subdifferential of the unpenalized objective function at the oracle estimator is helpful in understanding the behavior of $\partial \bar{\mathbf{k}}(\boldsymbol{\beta}, \boldsymbol{\xi})$. The oracle estimator for $\boldsymbol{\beta}_0^{(m)}$ is $\hat{\boldsymbol{\beta}}^{(m)} = \left(\hat{\boldsymbol{\beta}}_1^{(m)\prime}, \mathbf{0}_{p_n - q_n}'\right)'$ and across all quantiles is

$\hat{\hat{\boldsymbol{\beta}}} = \left(\hat{\boldsymbol{\beta}}^{(1)}, ..., \hat{\boldsymbol{\beta}}^{(M)}\right)$ and $\hat{\hat{\boldsymbol{\xi}}} = \left(\hat{\boldsymbol{\xi}}^{(1)}, ..., \hat{\boldsymbol{\xi}}^{(M)}\right)$. let $\hat{\hat{\boldsymbol{\beta}}}_j = \left(\hat{\beta}^{(1)}_{1j}, ..., \hat{\beta}^{(M)}_{1j}\right)'$ be the oracle estimator of $\bar{\beta}_j$. Define $\bar{\mathbf{s}}_j(\boldsymbol{\beta}, \boldsymbol{\xi}) = \left(s^{(1)}_j(\boldsymbol{\beta}, \boldsymbol{\xi}), ..., s^{(M)}_j(\boldsymbol{\beta}, \boldsymbol{\xi})\right)'$ as the subdifferential of the the unpenalized objective function, that is, the first part of (6.2) of the main paper, with respect to $\bar{\beta}_j$. The following lemma is an extension of Lemma 1 from the main paper.

Lemma B.6. *Assume Conditions B.1-B.6 are satisfied, $\lambda = o\left(n^{-(1-C_4)/2}\right)$, $n^{-1/2}\bar{q}_n = o(\lambda)$, $n^{-1/2}k_n = o(\lambda)$ and $\log(p_n) = o(n\lambda^2)$. There exists $a^{(m)*}_i$ with $a^{(m)*}_i = 0$ if $Y_i - \mathbf{x}_i'\hat{\boldsymbol{\beta}}^{(m)} - \boldsymbol{\Pi}(\mathbf{z}_i)'\hat{\boldsymbol{\xi}}^{(m)} \neq 0$ and $\mathbf{a}^{(m)*}_i \in [\tau - 1, \tau]$ otherwise. For $s^{(m)}_j(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\xi}})$ with $a_i = a^{(m)*}_i$, with probability approaching one,*

$$\bar{s}_j\left(\hat{\hat{\boldsymbol{\beta}}}, \hat{\hat{\boldsymbol{\xi}}}\right) = \mathbf{0}_M, \ j = 1, ..., \bar{q}_n \text{ or } j = p_n + 1, ..., p_n + L_n, \tag{B.10}$$

$$\left|\left|\hat{\hat{\beta}}_j\right|\right|_1 \geq (a + 1/2)\lambda, \ j = 1, ..., \bar{q}_n, \tag{B.11}$$

$$\left|\left|\bar{s}_j\left(\hat{\hat{\boldsymbol{\beta}}}, \hat{\hat{\boldsymbol{\xi}}}\right)\right|\right|_1 \leq c\lambda, \ \forall \ c > 0, \ j = \bar{q}_n + 1, ..., p_n. \tag{B.12}$$

Proof. Convex optimization theory provides that (B.10) holds. At least one element of $\bar{\beta}_{j0}$ is nonzero and thus (B.11) follows from (3.4) of Lemma 1 of the main paper. Lemma 1 of the main paper also implies (B.12). □

Let $\partial\bar{k}(\boldsymbol{\beta}, \boldsymbol{\xi})$ be the collection of $M(p_n + L_n)$-vectors that define the subdifferential of $\bar{k}(\boldsymbol{\beta}, \boldsymbol{\xi})$;

$$\partial\bar{k}(\boldsymbol{\beta}, \boldsymbol{\xi}) = \Big\{ (\boldsymbol{\kappa}_1', \boldsymbol{\kappa}_2', \ldots, \boldsymbol{\kappa}_{p_n+L_n}')' \in \mathbb{R}^{M(p_n+L_n)} : \\ \boldsymbol{\kappa}_j = \bar{\mathbf{s}}_j(\boldsymbol{\beta}, \boldsymbol{\xi}) + \lambda l_j, l_j \in \partial||\bar{\boldsymbol{\beta}}_j||_1, \ 1 \leq j \leq p_n; \\ \boldsymbol{\kappa}_j = \bar{\mathbf{s}}_j(\boldsymbol{\beta}, \boldsymbol{\xi}), \ p_n + 1 \leq j \leq p_n + L_n \Big\},$$

where $\partial||\bar{\boldsymbol{\beta}}_j||_1$ denotes the subdifferentiable of $||\bar{\boldsymbol{\beta}}_j||_1$ with respect to $\bar{\boldsymbol{\beta}}_j$. Let $\mathbf{b} = (b_1, \ldots, b_M)'$, then $\partial||\mathbf{b}||_1 = J_1 \times \cdots \times J_M$, where $J_k = [-1, 1]$ if $b_k = 0$; is equal to $\{1\}$ if $b_k > 0$ and is equal to $\{-1\}$ if $b_k < 0$; $1 \leq k \leq M$.

The subdifferential of $\bar{l}(\boldsymbol{\beta},\boldsymbol{\xi})$ is

$$\partial\bar{l}(\boldsymbol{\beta},\boldsymbol{\xi}) = \Big\{ (\boldsymbol{\mu}_1',\boldsymbol{\mu}_2',\ldots,\boldsymbol{\mu}_{p_n+L_n}')' \in \mathbb{R}^{M(p_n+L_n)} : \\ \boldsymbol{\mu}_j = l'(||\bar{\boldsymbol{\beta}}_j||_1)\partial||\bar{\boldsymbol{\beta}}_j||_1,\ 1 \le j \le p_n; \\ \boldsymbol{\mu}_j = \mathbf{0}_M,\ p_n+1 \le j \le p_n+L_n \Big\},$$

where l' is the ordinary derivative of $l(\cdot)$ given in Section 3.3 of the main paper.

Proof of Theorem 6.1

Proof. Proof of Theorem 6.1 follows the general structure of the proof of Theorem 3.1 and involves checking the condition of Lemma 7 of the main paper. However, unlike in the proof of Theorem 3.1, here $l(\boldsymbol{\beta},\boldsymbol{\xi})$ is convex but not differentiable everywhere due to the penalty being applied at the group level.

Define the set of $M(p_n+L_n)$-vectors:

$$\bar{\mathcal{G}} = \Big\{ \boldsymbol{\kappa} = (\boldsymbol{\kappa}_1,\boldsymbol{\kappa}_2,...,\boldsymbol{\kappa}_{p_n+J_n})' : \boldsymbol{\kappa}_j = \lambda\partial\left|\left|\hat{\bar{\boldsymbol{\beta}}}_j\right|\right|_1, j=1,..,\bar{q}_n; \\ \boldsymbol{\kappa}_j = \bar{\mathbf{s}}_j(\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\xi}}) + \lambda l_j,\ l_j \in \partial||\hat{\bar{\boldsymbol{\beta}}}_j||_1, j=\bar{q}_n+1,...,p_n; \\ \boldsymbol{\kappa}_j = \mathbf{0}_M, j = p_n+1,...,p_n+J_n, \Big\}$$

By Lemma B.6, we have $P\left(\bar{\mathcal{G}} \subset \partial k(\hat{\bar{\boldsymbol{\beta}}},\hat{\bar{\boldsymbol{\xi}}})\right) \to 1$.

Consider any $(\boldsymbol{\beta},\boldsymbol{\xi})'$ in a ball with the center $\left(\hat{\bar{\boldsymbol{\beta}}},\hat{\bar{\boldsymbol{\xi}}}\right)$ and radius $\lambda/2$. By Lemma 7 of the main paper, to prove the theorem it is sufficient to show that there exists $\boldsymbol{\kappa}^* = \left(\boldsymbol{\kappa}_1^{*'},\ldots,\boldsymbol{\kappa}_{p_n+J_n}^{*'}\right)' \in \mathcal{G}$ and $\boldsymbol{\mu}^* = (\boldsymbol{\mu}_1^{*'},\boldsymbol{\mu}_2^{*'},\ldots,\boldsymbol{\mu}_{p_n+L_n}^{*'})' \in \partial\bar{l}(\boldsymbol{\beta},\boldsymbol{\xi})$ such that

$$P\left(\boldsymbol{\kappa}_j^* = \boldsymbol{\mu}_j^*,\ j=1,...,p_n\right) \to 1, \tag{B.13}$$

$$P\left(\boldsymbol{\kappa}_{p_n+j}^* = \boldsymbol{\mu}_{p_n+j}^*,\ j=1,...,J_n\right) \to 1, \tag{B.14}$$

as $n,p \to \infty$.

Since $\forall\ \mathbf{u} = (\boldsymbol{\mu}_1',\boldsymbol{\mu}_2',\ldots,\boldsymbol{\mu}_{p_n+L_n}')' \in \partial\bar{l}(\boldsymbol{\beta},\boldsymbol{\xi})$ we have $\mathbf{u}_{p_n+j} = \mathbf{0}_M$, for $j=1,...,J_n$, (B.14) is satisfied for any $\mathbf{u} \in \partial\bar{l}(\boldsymbol{\beta},\boldsymbol{\xi})$ and any $\boldsymbol{\kappa} \in \mathcal{G}$. In the following, we outline how we can find $\boldsymbol{\kappa}^* \in \mathcal{G}$ and $\boldsymbol{\mu}^* \in \partial \bar{l}(\boldsymbol{\beta}, \boldsymbol{\xi})$ so that (B.13) is satisfied.

1. For $1 \le j \le \bar{q}_n$, we have $||\bar{\boldsymbol{\beta}}_j^0||_1 > 0$.
   By Lemma B.6, we have
   $$\begin{aligned} \min_{1\le j\le \bar{q}_n} ||\bar{\boldsymbol{\beta}}_j||_1 &\ge \min_{1\le j\le \bar{q}_n} ||\hat{\bar{\boldsymbol{\beta}}}_j||_1 - \max_{1\le j\le \bar{q}_n} \left|\left|\hat{\bar{\boldsymbol{\beta}}}_j - \bar{\boldsymbol{\beta}}_j\right|\right|_1 \\ &\ge (a+1/2)\lambda - \lambda/2 = a\lambda, \end{aligned}$$
   with probability approaching one. Hence, with probability approaching one, $l'(||\bar{\boldsymbol{\beta}}_j||_1) = \lambda$ for either SCAD or MCP penalty function; and $\boldsymbol{\mu}_j = \lambda \mathbf{c}_j$ for some $\mathbf{c}_j \in \partial ||\bar{\boldsymbol{\beta}}_j||_1$. For any $1 \le j \le \bar{q}_n$, $\left|\left|\hat{\bar{\boldsymbol{\beta}}}_j - \bar{\boldsymbol{\beta}}_{0j}\right|\right|_1 = O_p\left(n^{-1/2}\bar{q}_n^{1/2}\right) = o_p(\lambda)$. Recall $\bar{\boldsymbol{\beta}}_j$ is within radius $\lambda/2$ of $\hat{\bar{\boldsymbol{\beta}}}_j$. Then with probability approaching one, the entries of $\hat{\bar{\boldsymbol{\beta}}}_j$ and $\bar{\boldsymbol{\beta}}_j$ that correspond to the nonzero components of $\bar{\boldsymbol{\beta}}_{0j}$ share the same signs. For any $\boldsymbol{\kappa} \in \mathcal{G}$, we can write $\boldsymbol{\kappa}_j = \lambda \mathbf{d}_j$ for some $\mathbf{d}_j \in \partial ||\hat{\bar{\beta}}_j||_1$. Recall that for $\mathbf{b} = (b_1, \ldots, b_M)'$, we have $\partial ||\mathbf{b}||_1 = J_1 \times \cdots \times J_M$, where $J_k = [-1, 1]$ if $b_k = 0$; is equal to $\{1\}$ if $b_k > 0$ and is equal to $\{-1\}$ if $b_k < 0$; $1 \le k \le M$. Hence, we can find $\mathbf{c}_j^* \in \partial ||\bar{\boldsymbol{\beta}}_j||_1$ and $\mathbf{d}_j^* \in \partial ||\hat{\bar{\beta}}_j||_1$ such that $\mathbf{c}_j^* = \mathbf{d}_j^*$. Let $\boldsymbol{\mu}_j^* = \lambda \mathbf{c}_j^*$ and $\boldsymbol{\kappa}_j^* = \lambda \mathbf{d}_j^*$, $1 \le j \le \bar{q}_n$.
2. For $j = \bar{q}_n + 1, ..., p_n$, $\hat{\bar{\boldsymbol{\beta}}}_j = \mathbf{0}_M$ by the definition of the oracle estimator. Note that
   $$||\bar{\boldsymbol{\beta}}_j||_1 \le ||\hat{\bar{\boldsymbol{\beta}}}_j| + ||\hat{\bar{\boldsymbol{\beta}}}_j - \bar{\boldsymbol{\beta}}_j||_1 < \lambda/2.$$
   Hence, for the SCAD penalty function $l'(||\bar{\boldsymbol{\beta}}_j||_1) = \mathbf{0}_M$ and for the MCP penalty function $l'(||\bar{\boldsymbol{\beta}}_j||_1) = a^{-1}||\bar{\beta}_j||_1$. Thus, for the SCAD penalty function $\boldsymbol{\mu}_j = \mathbf{0}_M$; for the MCP penalty function $\boldsymbol{\mu}_j = a^{-1}||\bar{\beta}_j||_1 \mathbf{c}_j$ for some $\mathbf{c}_j \in \partial ||\bar{\boldsymbol{\beta}}_j||_1$; $j = \bar{q}_n + 1, ..., p_n$. Note that $a^{-1}||\bar{\beta}_j||_1 \le \lambda/2$.
   For any $\boldsymbol{\kappa} \in \mathcal{G}$, we can write $\boldsymbol{\kappa}_j = \bar{\mathbf{s}}_j(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\xi}}) + \lambda \mathbf{d}_j$, $\mathbf{d}_j \in \partial ||\hat{\bar{\boldsymbol{\beta}}}_j||_1 = [-1, 1] \times \ldots \times [-1, 1]$. For both penalty functions, by lemma B.6 we have $||\bar{\mathbf{s}}_j(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\xi}})||_1 \le \lambda/2$, $j = \bar{q}_n + 1, ..., p_n$ with probability approaching one. For the SCAD penalty function, we can find $\mathbf{d}_j^* \in \partial ||\hat{\bar{\boldsymbol{\beta}}}_j||_1$ such that $\bar{\mathbf{s}}_j(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\xi}}) + \lambda \mathbf{d}_j^* = 0$. For the MCP penalty function and for any $\mathbf{c}_j^* \in \partial ||\bar{\boldsymbol{\beta}}_j||_1$, we can find $\mathbf{d}_j^* \in \partial ||\hat{\bar{\boldsymbol{\beta}}}_j||_1$ such that $\bar{\mathbf{s}}_j(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\xi}}) + \lambda \mathbf{d}_j^* = a^{-1}||\bar{\beta}_j||_1 \mathbf{c}_j^*$ with probability approaching one. Let $\boldsymbol{\mu}_j^* = 0$ for SCAD penalty function and $\boldsymbol{\mu}_j^* = \lambda \mathbf{c}_j^*$ for MCP penalty function; let $\boldsymbol{\kappa}_j^* = \bar{\mathbf{s}}_j(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\xi}}) + \lambda \mathbf{d}_j^*$, $j = \bar{q}_n + 1, ..., p_n$.

This finishes the proof. $\square$

REFERENCES

[1] Breheny, P. and Lee, S. (2015). ncvreg: Regularization Paths for SCAD and MCP Penalized Regression Models. R package version 3.4-0. Available at http://CRAN.R-project.org/package=ncvreg.

[2] de Boor, C. (1976). Splines as linear combinations of B-splines. A Survey. G.G. Lorentz, C.K. Chui, L.L. Schumaker (Eds.), *Approximation Theory II*, 1–47, Academic Press, New York.

[3] Koenker, R. (2005). *Quantile Regression*, Cambridge University Press.

[4] Robinson, P. (1988). Root-N-Consistent Semiparametric Regression. *Econometrica*, **56**, 931-954.

[5] Schumaker, L. (1981). *Spline Functions: Basic Theory*, Wiley: New York.

[6] Shi,P. and Li, G. (1995). Global convergence rates of B-spline M-estimators in nonparametric regression. *Statistica Sinica*, **5**, 303-318.

[7] Stone, C. (1985). Additive regression and other nonparametric models. *Annals of Statistics*, **13**, 689-706.

[8] Wang, L., Wu, Y. and Li, R. (2012). Quantile regression for analyzing heterogeneity in ultra-high dimension. *Journal of the American Statistical Association*, **107**, 214-222.

[9] Xue, L. and Yang, L. J. (2006). Additive coefficient modeling via polynomial spline. *Statistica Sinica*, **16**, 1423-1446.

B. Sherwood
Department of Biostatistics
Johns Hopkins University
Baltimore, MD 21205, USA
E-mail: bsherwo2@jhu.edu

L. Wang
School of Statistics
University of Minnesota
Minneapolis, MN 55455, USA
E-mail: wangx346@umn.edu